

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/159641>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

# **Context-Aware Support for Stress Self-Management: From Theory to Practice**

Proefschrift

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,  
volgens besluit van het college van decanen  
in het openbaar te verdedigen op maandag 10 oktober 2016  
om 14:30 uur precies

door

Saskia Julia Koldijk

geboren op 7 mei 1987  
te Emmerich (Duitsland)

Promotoren:

Prof. dr. Wessel Kraaij

Prof. dr. Mark A. Neerincx (Technische Universiteit Delft)

Manuscriptcommissie:

Prof. dr. ir. Theo P. van der Weide

Prof. dr. Wijnand IJsselsteijn (Technische Universiteit Eindhoven)

Prof. dr. ir. Hermie Hermens (Universiteit Twente)

Prof. dr. ir. Judith Masthoff (University of Aberdeen, Verenigd Koninkrijk)

Prof. dr. Daniel Gatica-Perez (EPFL, Lausanne, Zwitserland)



SIKS Dissertation Series No. 2016-43

The research reported in this thesis has been carried out under the auspices of the Dutch Research School for Information and Knowledge Systems (SIKS), the institute for Computing and Information Science (iCIS) of the Radboud University Nijmegen, and TNO.

# COMMIT/

This research was supported by the Dutch national program COMMIT (project P7 SWELL).

Copyright © 2016 Saskia Koldijk

Cover design: Karli Grafisch [www.karligrafisch.nl](http://www.karligrafisch.nl)

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress

Published by: Uitgeverij BOXPress, 's-Hertogenbosch

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation: Support technology to manage stress at work . . . . .	2
1.2	Approach . . . . .	4
1.3	Theory: Work stress and interventions . . . . .	5
1.4	Technology: Inferring working context and mental states . . . . .	6
1.5	Users: Needs and concerns . . . . .	8
1.6	My publications . . . . .	9
<b>2</b>	<b>Deriving requirements from work stress and intervention theory</b>	<b>11</b>
2.1	Introduction . . . . .	12
2.2	Initial study on user needs . . . . .	15
2.3	Well-being at work concepts . . . . .	16
2.3.1	Engagement and burn-out . . . . .	16
2.3.2	Stress . . . . .	17
2.3.3	Conclusions: Relevant concepts for the system . . . . .	19
2.4	Causes of work stress . . . . .	20
2.4.1	Job Demands-Resources Model . . . . .	21
2.4.2	Effort-Reward Imbalance Model . . . . .	24
2.4.3	Effort-Recovery Model . . . . .	24
2.4.4	Person-Environment Fit Model . . . . .	26
2.4.5	Conclusions: Addressing causes of work stress . . . . .	27
2.5	Inferring stress and its context . . . . .	27
2.5.1	Technical possibilities . . . . .	30
2.5.2	User choices regarding data collection . . . . .	32
2.5.3	Conclusions: Using sensing and reasoning . . . . .	32
2.6	Improving well-being at work . . . . .	33
2.6.1	Intervention theory . . . . .	34
2.6.2	Behavioral change . . . . .	34
2.6.3	Conclusions: Technology based interventions . . . . .	36

2.7	Envisioned system and evaluation of prototypes . . . . .	38
2.7.1	Insight in stress sources - SWELL Workload Mirror . . . . .	39
2.7.2	Fostering colleague support - SWELL Fishualization . . . . .	42
2.7.3	Providing tips - SWELL NiceWork e-coach . . . . .	45
2.7.4	Providing support - SWELL HappyWorker system . . . . .	48
2.7.5	Conclusions: Evaluation of prototypes . . . . .	50
2.8	Conclusions and Discussion . . . . .	51
2.8.1	Conclusions . . . . .	51
2.8.2	Discussion . . . . .	52
2.8.3	Identified research challenges and opportunities . . . . .	54
<b>3</b>	<b>The SWELL knowledge work dataset</b>	<b>57</b>
3.1	Introduction . . . . .	58
3.2	Related Work . . . . .	60
3.3	Data Collection Context . . . . .	62
3.3.1	Design . . . . .	62
3.3.2	Tasks . . . . .	63
3.3.3	Procedure . . . . .	64
3.3.4	Apparatus . . . . .	64
3.3.5	Subjective Ratings . . . . .	64
3.3.6	Sensors . . . . .	65
3.3.7	Participants . . . . .	66
3.4	Dataset . . . . .	68
3.4.1	The Feature Data . . . . .	68
3.4.2	The Raw Data and Preprocessing . . . . .	68
3.5	Example Analyses . . . . .	71
3.6	Discussion . . . . .	73
3.7	Conclusion . . . . .	74
<b>4</b>	<b>Real-time task recognition based on computer activities</b>	<b>75</b>
4.1	Introduction . . . . .	76
4.2	Task recognition framework . . . . .	77
4.2.1	Task Labels . . . . .	77
4.2.2	Features . . . . .	78
4.2.3	Classifiers . . . . .	79
4.3	Approach for framework evaluation . . . . .	79
4.3.1	Tool For Collecting Annotated Data . . . . .	80
4.3.2	Method . . . . .	80
4.4	Analyses and results . . . . .	83
4.4.1	Task Labels and Features . . . . .	83

4.4.2	Comparison of Classifiers . . . . .	83
4.4.3	Individual Differences . . . . .	85
4.5	Discussion . . . . .	87
4.6	Conclusions . . . . .	89
<b>5</b>	<b>Visual analytics of work behavior data</b>	<b>91</b>
5.1	Introduction . . . . .	92
5.2	Related work . . . . .	93
5.2.1	Affective computing . . . . .	93
5.2.2	Visual analytics . . . . .	93
5.3	Data analysis and visual analytics . . . . .	94
5.3.1	General overview . . . . .	94
5.3.2	Relating subjective experience to sensor data . . . . .	97
5.3.3	Typical user behavior groups . . . . .	97
5.3.4	Filtering the set of features . . . . .	98
5.3.5	Exploration of facial activity patterns . . . . .	98
5.3.6	Details on demand: Visualizing facial activity . . . . .	98
5.3.7	Grouping typical facial expressions . . . . .	99
5.4	Conclusion and future work . . . . .	100
<b>6</b>	<b>Detecting work stress in offices by combining unobtrusive sensors</b>	<b>103</b>
6.1	Introduction . . . . .	104
6.2	Related work . . . . .	106
6.3	Dataset . . . . .	109
6.4	Using several unobtrusive sensors to detect stress in office environments . . . . .	111
6.4.1	Inferring the working condition . . . . .	112
6.4.2	Predicting mental states . . . . .	115
6.5	Taking into account individual differences . . . . .	120
6.5.1	Individual differences regarding working condition . . . . .	121
6.5.2	Individual differences regarding mental states . . . . .	122
6.5.3	Addressing individual differences . . . . .	123
6.6	Conclusions and Discussion . . . . .	125
<b>7</b>	<b>Human-centered development of context aware support systems</b>	<b>129</b>
7.1	Introduction . . . . .	130
7.2	Theoretical and technological foundation . . . . .	135
7.2.1	Derive core functions from human factors . . . . .	135
7.2.2	Functional components from enabling technology . . . . .	137
7.3	User and operational support demands . . . . .	141

7.3.1	Identify user needs . . . . .	141
7.3.2	Identify user concerns . . . . .	142
7.3.3	Address privacy . . . . .	145
7.4	Harmonizing all identified requirements . . . . .	151
7.5	Evaluation for further refinement . . . . .	153
7.5.1	Method (user study 3) . . . . .	153
7.5.2	Results . . . . .	155
7.5.3	Conclusions and discussion . . . . .	158
7.5.4	Refinement of the user model . . . . .	161
7.6	Conclusions and Discussion . . . . .	162
7.6.1	Human-centered development method . . . . .	162
7.6.2	User studies . . . . .	164
<b>8</b>	<b>Privacy and user trust in context-aware systems</b>	<b>167</b>
8.1	Introduction . . . . .	168
8.2	Context aware system use-case: SWELL . . . . .	169
8.3	Privacy aspects . . . . .	170
8.4	Privacy by Design . . . . .	172
8.5	User perspective: Evaluation study . . . . .	174
8.5.1	Method . . . . .	174
8.5.2	Results . . . . .	177
8.6	Discussion . . . . .	179
8.7	Conclusions . . . . .	180
<b>9</b>	<b>Conclusions and reflection</b>	<b>181</b>
9.1	Theory: Work stress and interventions . . . . .	181
9.2	Technology: Inferring working context and mental states . . . . .	184
9.3	Users: Needs and concerns . . . . .	186
9.4	Reflection . . . . .	188
9.4.1	Addressed research challenges and limitations . . . . .	188
9.4.2	Support technology to manage stress at work . . . . .	189
	<b>Bibliography</b>	<b>191</b>
	<b>Summary</b>	<b>209</b>
	<b>Samenvatting</b>	<b>213</b>
	<b>Zusammenfassung</b>	<b>217</b>
	<b>Acknowledgements</b>	<b>221</b>

---

# Introduction

Recent technological advances, like sensors available in smart phones and e-coaching apps, offer radically new methods for support systems. This research aims at investigating how such new technologies can be used to support people to cope better with stress at work, to prevent burn-out. The EU Horizon2020 call on ‘Health, Demographic Change and Wellbeing’ ([Horizon2020 2015](#)) highlights the importance of personalized health, in particular aiming to ‘improve our ability to monitor health and to prevent, detect, treat and manage disease’.

A successful implementation of such a new support system for well-being at work, however, poses challenges in several research areas:

1. Behavioral/Social Sciences and Occupational Health: To build an *effective* support system, relevant existing work stress theories should be applied. Traditionally, social sciences collect data using questionnaires. New sensors and e-coaching technologies, however, provide interesting alternative means to collect data and provide interventions. The challenge is to extend, refine and formalize existing theories in such a way, that they can incorporate new data acquisition and processing technology.
2. Computer Science and Artificial Intelligence: To build an *innovative* support system, the possibilities of new technologies should be investigated. We are in particular interested in the potential of different sensors to infer relevant aspects of the user’s current context. Collecting data with sensors is rather simple, the challenge is to make sense of this low level data, e.g. in terms of the tasks worked on, experienced stress or mental effort.
3. User-oriented Design and Requirements Engineering: To build a *user-friendly* support system, the needs and concerns of users should be taken into account. A system that uses sensors to collect data may in particular



pose privacy concerns that need to be addressed. The challenge is to harmonize functional requirements and privacy requirements in the design of context aware systems.

We address the challenges from these three different research fields. Therefore, this thesis contributes to a better understanding of how theoretical (work stress) models can be operationalized, and how sensors can be used to infer context aspects (relevant to work stress). Moreover, it contributes to a better understanding of how theoretical and technological insights can be combined with input on user needs and concerns, to build effective, innovative and user-friendly support systems.

## **1.1 Motivation: Support technology to manage stress at work**

The support system that we aim to develop addresses the problem of work stress. We probably all know stressful working days. Imagine Bob, a typical knowledge worker, who is predominantly concerned with interpreting and generating information.

Bob gets into the office at 9, starts up his computer, takes a look at his emails and calendar and plans what things he has to do this day. Then he starts working on one of the important tasks that have to be completed this week. When an e-mail comes in, he quickly reads it. As it is not relevant to him, he continues his task. When a colleague drops by, Bob looks up and talks to him, which results in some more to-do's for the day. As they are quite urgent, Bob decides to do them right away. After completion he switches back to his previous task and continues his working day. At 5 'o clock Bob notices that he has not completed all planned tasks yet and he feels somewhat stressed. He starts wondering where the time went and whether he should work overtime again to finish up.

Bob and many other knowledge workers experience typical working days like this, which can easily cause stress (Michie, 2002). The question is: When does stress become a danger to well-being? And how can we help people to handle stress appropriately?

To date, the problem of work stress is often approached with questionnaires in which employees are asked to rate various aspects of their work (e.g. Zwi-eten *et al.*, 2014; Kraan *et al.*, 2000), followed by department wide interventions, such as redesigning the work or providing training. As knowledge workers are relatively flexible in their work (when they do what and how they work),

there is great potential for them to contribute to the improvement of their own well-being.

New technologies are emerging, such as new sensors, smart reasoning and e-coaching apps. Trends like ‘quantified self’ (see e.g. <http://quantifiedself.com>) already show a potential of collecting personal sensor data (e.g. heart rate, activity patterns) for health improvement. In their paper on technology for well-being, IJsselsteijn *et al.* (2006) describe that advancements in sensing and interpretation are promising. They further state that using technology for improving well-being has many advantages, e.g. its persistence or objectiveness, the possibility to provide just-in-time notifications with relevant, actionable information or their supportive and motivating role.

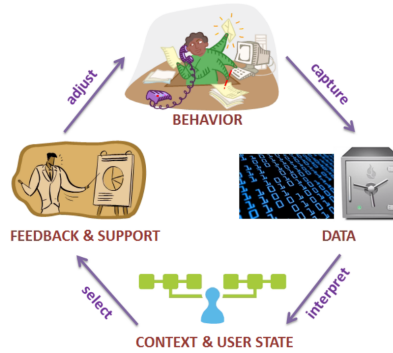


Figure 1.1: SWELL approach.

In the SWELL project (Smart Reasoning for Well-being at Home and at Work)<sup>1</sup> we explore how new pervasive context aware systems (CAS) can address well-being at work at an individual level. With ‘pervasive’ we mean, that the systems are integrated into the day to day working scenario. With ‘context aware’ we mean, that the system takes into account information about the current context of the user. Figure 1.1 depicts our approach. We see possibilities in using unobtrusive and easily available sensors to capture the knowledge workers behavior (e.g. computer interactions, webcam for facial expressions, Kinect for postures) and infer the current working context and mental state. Based upon this information, we aim to develop supporting technology that is context-aware, so optimally adapted to the situation and state of the user. Knowledge workers can then directly act, gaining a more healthy work style and preventing stress building up.

<sup>1</sup><http://www.swell-project.net>

## 1.2 Approach

To develop a, theoretically and empirically grounded, context aware support system, we take a multi-disciplinary approach. We apply the situated Cognitive Engineering (sCE) method (Neerincx and Lindenberg, 2008) and combine theory on work stress with technological possibilities, taking in mind input on user needs. The remainder of this thesis is therefore divided into three parts: 1) Theory, 2) Technology, and 3) Users.

1. **Theory** provides a foundation for the design of CAS and yields claims, e.g. what effect a specific functionality would have.
2. **Technology** yields core functions for the system, e.g. sensors are used to infer context information.
3. **User demands** provide use cases, i.e. the context in which the final system is used, like the scenario of Bob.

Claims, core functions and use cases together describe the system specification and provide requirements for the system. Note that sCE does not prescribe a specific order in time; the sCE-processes are iterative and can partially take place in parallel. A first specification of the system can be evaluated with users and then be further refined.

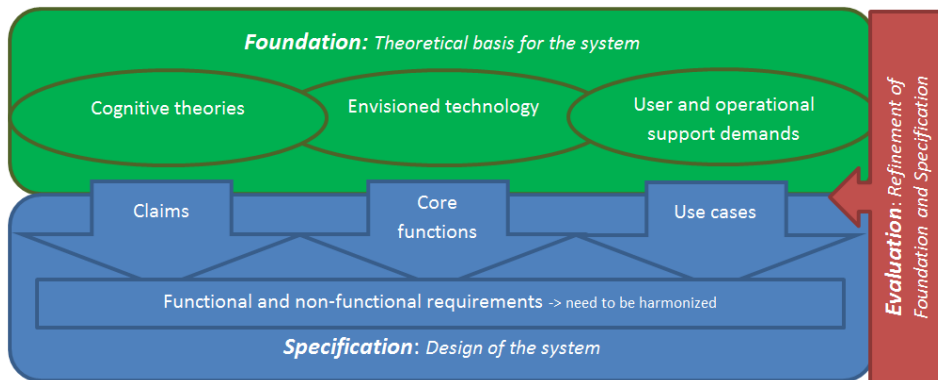


Figure 1.2: Situated Cognitive Engineering (sCE) approach

### 1.3 Theory: Work stress and interventions

The general objective of the SWELL system is to improve well-being at work. In the first part of our research (Chapter 2), we delve into work stress and intervention theory to build an understanding of the problem of work stress and find determinants to address.

Different disciplines view the world from different angles: work psychology describes high level principles (e.g. demands vs. resources); biology describes low level physiological processes (e.g. changes in heart rate); and behavioral psychology describes determinants to address for behavioral change (e.g. motivation). New technologies bring new possibilities. With sensors, we can do measurements in real-time, e.g. monitor computer activities or changes in heart rate. With pervasive technology we can enable support that is context aware, i.e. optimally adapted to the situation and state of the user. We think that psychology can benefit from utilizing new technological possibilities, whereas system designers can benefit from using existing theories. To really do multidisciplinary work like this, it is necessary that the experts understand each others domains well, which is challenging.

In Chapter 2, we take up the challenge to relate theories from different fields to each other (i.e. work stress, interventions and behavioral change) and provide a unified overall framework regarding well-being at work, that serves as a theoretical basis. We aim to answer the following research questions: **RQ1** *‘Which concepts are relevant with respect to well-being at work?’* and **RQ2** *‘Which person, work and context conditions can lead to negative stress?’*. In our work, we aim to integrate new technological possibilities into the theoretical framework. We answer the following questions: **RQ3** *‘How can sensors be applied to automatically infer stress and the context in which it appears?’* and **RQ4** *‘Which interventions can be provided by means of pervasive technology to help a knowledge worker improve his well-being at work?’*. Our framework aims to bring the fields of psychology and software engineering closer together to address well-being at work in an innovative way. Based upon this framework, we formulate requirements for the envisioned system for improving well-being at work. We present four resulting prototypes, and first evaluation studies with potential end-users.

## 1.4 Technology: Inferring working context and mental states

In the second part of our research, we investigate computational approaches to inferring the user's current working context and mental state from unobtrusive sensors (Chapters 3, 4, 5 and 6).

To be able to test our algorithms for context and mental state detection, a rich dataset is required. In related work, we find several drawbacks: Work in the area of affective computing investigates the possibility of inferring stress and emotion from sensor data. We found that often rather sophisticated, expensive and/or obtrusive equipment is used (e.g. eye tracker, body sensors) in controlled lab settings. Typically standardized tasks are used, such as remembering digits to manipulate mental effort, or watching emotional movie excerpts to influence emotions. These tasks are not representative of 'real' office work. Work on user state modeling is often performed in a process control context, e.g. on naval ships or in flight control. Only little work is done on user state modeling in an office context.

**Dataset.** In Chapter 3, we describe how we created a dataset, that overcomes drawbacks that are typically observed in related work: Instead of a rather artificial task, we used a realistic office setting while stressors were manipulated systematically (time pressure, email interruptions). Instead of expensive and/or obtrusive equipment, we use a varied set of (unobtrusive and easily available) sensors: computer logging, video, Kinect 3D and body sensors. Finally, instead of only collecting data for our own use, the dataset is made available to the scientific community, in raw and already preprocessed form.

**Task recognition.** By looking back at their own working behavior, knowledge workers might get a better grip on their work style and improve it. Some commercial off-the-shelf computer applications present an overview of time spent per application and websites browsed (e.g. <http://www.manictime.com/>), but these require the user to interpret for which task a specific program or website was used. Minimal effort should be required from the user. We aim at automatic recognition of tasks based on computer activities. In related work on user activity recognition, activities with rather clear structures and predefined steps are being modeled (e.g. computer activities such as form filling or planning a meeting (Rath, Devaurs, and Lindstaedt, 2009); user's goals in a computer game (Albrecht *et al.*, 1997)). Therefore, often model-based classification is applied, e.g. modeling the sequence of actions in time.

In Chapter 4, we investigate the following research question: **RQ5** '*Can knowledge workers' tasks be recognized based upon computer interactions?*' These tasks are

less structured and sequences are more spontaneous. We apply different machine learning approaches to real-world office data.

**Visual analytics.** Ideally, we want to provide knowledge workers insight in their work behavior and how this relates to their well-being. In general, the affective computing community often uses black-box machine learning algorithms to classify sensor data directly into mental states (see Sharma and Gedeon (2012)). The focus is most often on generating one model, that generalizes well to new users. We are, however, more interested in behavioral patterns of individual users, that are related to stress. Visualizing such patterns may give users more insight and actionable information than just a stress labeling.

In Chapter 5 we address the research question: **RQ6** *‘How can sensor data be used to gain insight into work behavior, specifically related to stress at work?’*. We apply visual analytics to the SWELL-KW dataset. In an iterative approach, we combine automatic data analysis procedures with visualization techniques, to gain deeper insight into the data. We specifically focus on differences between users and how to cope with them in data processing and visualization. The final aim is to develop a visualization system for individual users that gives insight in a large amount of behavioral data recorded with sensors.

**Detecting stress.** We are interested in detecting work stress. We identified two methodological and applied machine learning challenges: a) In related work often sophisticated sensors (e.g. eye tracker, body sensors) are used in controlled lab settings; and b) often one general model is learned over all users, which is currently the typical statistical modeling approach for e.g. emotion or stress recognition. However, we found that people differ in their (work) behavior: normal behavior of users already differs. The way in which people express mental effort or stress may also differ.

In Chapter 6 we therefore aim to address these challenges. We see possibilities to build human state estimation techniques by combining information from multiple weak indicator variables based on physically unobtrusive measurements. We answer our research question: **RQ7** *‘Can we distinguish stressful from non-stressful working conditions, and can we estimate mental states of office workers by using several unobtrusive sensors?’* We present how well several machine learning approaches perform on this task. We also present the most useful modalities and features. The other research question we answer is: **RQ8** *‘How important are individual differences?’* Especially for estimating mental states, individual differences seem to play an important role, so we decided to build models for particular user groups.

## 1.5 Users: Needs and concerns

In the third part of our research, we delve into human-centered design, for developing effective and user-friendly context aware support systems (Chapters 7 and 8).

To provide contextualized and personalized support, context aware systems need to collect personal data. Tension may arise between technological possibilities of building rich user models and concerns that users might have, in particular regarding privacy. However, a compact and coherent development method, that balances functionality and privacy needs, is lacking. The main focus is investigating how a system applying user modeling can be designed in a way, that takes the user's privacy concerns into account.

In Chapter 7 we address the following research question: **RQ9** *'How can we refine the 'situated cognitive engineering' methodology on two aspects: a) defining the context during the requirements engineering process, and b) addressing functional and non-functional requirements coherently?'* The development of the SWELL system provides an interesting use-case with challenges around personal data collections. We particularly focus on analyzing user concerns, complementing the analysis with a privacy impact assessment, and suggesting ways to address privacy in CAS.

In Chapter 8, we focus on privacy and user trust in CAS. We first answer the research question: **RQ10** *'How should privacy be addressed in the design of CAS?'* We then present a user study, in which we address the research question: **RQ11** *'What effect does information on privacy by design have on users?'* We test two hypotheses: a) *'When users have access to detailed information on data collection and privacy by design, they have less privacy concerns and more trust in the system.'* b) *'Users have a more positive attitude towards using the CAS, as a consequence of increased trust in the system.'*

## 1.6 My publications

- Koldijk, S., Kraaij, W., & Neerincx, M.A. **Deriving Requirements for Pervasive Well-Being Technology From Work Stress and Intervention Theory: Framework and Case Study.** *Journal of Medical Internet Research: mHealth and uHealth*, 2016.
- Koldijk, S., Neerincx, M.A., & Kraaij, W. **Detecting work stress in offices by combining unobtrusive sensors.** *IEEE Transactions on Affective Computing*, submitted.
- Koldijk, S., Siljee, J., Neerincx, M.A., & Kraaij, W. **Human-Centered Development Method for Effective and Privacy-Friendly Context Aware Support Systems - A Case Study.** *Cognition, Technology & Work*, submitted.
- Koldijk, S., Bernard, J., Ruppert, T., Kohlhammer, J., Neerincx, M.A., & Kraaij, W. (2015). **Visual Analytics of Work Behavior Data - Insights on Individual Differences.** In: *Proceedings of the EG/VGTC Conference on Visualization (EuroVis 2015)* (Cagliari, Italy, 25-29 May 2015).
- Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M.A., & Kraaij, W. (2014). **The SWELL Knowledge Work Dataset for Stress and User Modeling Research.** In: *Proceedings of the 16th ACM International Conference on Multimodal Interaction (ICMI 2014)* (Istanbul, Turkey, 12-16 November 2014).
- Schavemaker, J., Boertjes, E., Koldijk, S., Wiertz, L., Verberne, S., Sappelli, M., & Kaptein, R. (2014). **Fishualization: a group feedback display.** In: *Proceedings of the 9th International Conference on Methods and Techniques in Behavioral Research (Measuring Behavior 2014)* (Wageningen, The Netherlands, 27-29 August 2014).
- Koldijk, S., Koot, G., Neerincx, M.A., & Kraaij, W. (2014). **Privacy and User Trust in Context-Aware Systems.** In: *Proceedings of the 22nd Conference on User Modeling, Adaptation and Personalization (UMAP 2014)* (Aalborg, Denmark, 7-11 July 2014).
- Koldijk, S., Neerincx, M.A., & Kraaij, W. (2014). **Using sensors in the office to support knowledge workers with self-management of well-being.** In: *Proceedings of Well-being at Work 2014* (Copenhagen, Denmark, 26-28 May 2014).
- Sappelli, M., Verberne, S., Koldijk, S., & Kraaij, W. (2014). **Collecting a dataset of information behaviour in context.** In: *Proceedings of the 4th*



*Workshop on Context-awareness in Retrieval and Recommendation (CARR @ ECIR 2014)* (Amsterdam, The Netherlands, 13-16 April 2014).

- Koldijk, S., Sappelli, M., Neerincx, M.A., & Kraaij, W. (2013). **Unobtrusive monitoring of knowledge workers for stress self-regulation**. In: *Proceedings of the 21nd Conference on User Modeling, Adaptation and Personalization (UMAP 2013)* (Rome, Italy, 10-14 Juli 2013).
- Verberne, S., van der Heijden, M., Hinne, M., Sappelli, M., Hoenkamp, E., Koldijk, S., & Kraaij, W. (2013). **Reliability and Validity of Query Intent Assessments**. *Journal of the American Society for Information Science and Technology*, Volume 64, Issue 11, pages 2224–2237.
- Koldijk, S., Neerincx, M.A., & Kraaij, W. (2012). **Unobtrusively measuring stress and workload of knowledge workers**. In: *Proceedings of Measuring Behavior 2012* (Utrecht, The Netherlands, 28-31 August 2012).
- Koldijk, S., van Staalduinen, M., Neerincx, M.A., & Kraaij, W. (2012). **Real-time task recognition based on knowledge workers' computer activities**. In: *Proceedings of the European Conference on Cognitive Ergonomics (ECCE 2012)* (Edinburgh, Scotland, 28-31 August 2012).
- Koldijk, S. (2012). **Automatic recognition of context and stress to support knowledge workers**. In: *Proceedings of the European Conference on Cognitive Ergonomics (ECCE 2012)* (Edinburgh, Scotland, 28-31 August 2012; doctoral consortium).
- Koldijk, S. (2012). **Capturing context and mental state of knowledge workers**. In: *Proceedings of the Fourth Information Interaction in Context Conference (IiX 2012)* (Nijmegen, The Netherlands, 21-24 August 2012; doctoral consortium).

# Deriving requirements for pervasive well-being technology from work stress and intervention theory: Framework and case study

*Ideally, new support systems are grounded in existing work stress and intervention theory. However, there is a large diversity of theories and these theories do hardly provide explicit directions for technology design. In this chapter, we present a comprehensive and concise framework that can be used to design pervasive technologies that support knowledge workers to decrease stress. Based on a literature study, we 1) identify concepts relevant to well-being at work, and 2) select different work stress models to find causes of work stress that can be addressed. From a technical perspective, we then 3) describe how sensors can be used to infer stress and the context in which it appears. Finally, we 4) use intervention and behavioral change theory to further specify interventions that can be provided by means of pervasive technology. The resulting framework combines theories on work stress and interventions, and integrates technological possibilities of sensing and support. We used this framework to derive requirements for pervasive well-being technology, and present 4 prototypes that were implemented. Evaluation studies show that potential end users are positive. Finally, we provide 7 key research challenges that were identified in the area of pervasive systems for well-being.*

---

This chapter is based on Koldijk, Kraaij, and Neerinx (2016). "Deriving requirements for pervasive well-being technology from work stress and intervention theory: Framework and case study". Published in: Journal of Medical Internet Research: mHealth and uHealth.

## 2.1 Introduction

Employees often report the experience of stress at work, which is related to their well-being. In this research we focus on the population of knowledge workers, who are predominantly concerned with interpreting and generating information. Stress is easily caused by their typical working conditions (Michie, 2002). Several tasks that need to be finished before a deadline, and their course of action is not always self-planned but also determined by external causes, like phone calls, mails, information requests, other persons or appointments (Czerwinski, Horvitz, and Wilhite, 2004a). According to Demerouti, A. B. Bakker, Nachreiner, *et al.* (2001) the feeling of demands outgrowing personal resources causes stress. Employees may fall short to handle adequately with stress (coping), or may fail to relax or detach from work in the evenings, impairing recovery. When stress builds up this can be a danger to well-being, in the worst case resulting in burn-out. The employee should recognize when stress becomes problematic. Stress can either directly lead to illness through its physiological effects or indirectly, through maladaptive health behavior, such as smoking, poor eating habits or lack of sleep (J. Bakker *et al.*, 2012). Following the definition by Selye (1956), an employee complaining about stress might thus mean that his working conditions are very demanding (the stressor), or that he feels that demands put upon him are higher than he can take (the perception of stressors) or that he feels stress reactions in his body (the experience of stress). To date, the problem of work stress is often approached with questionnaires in which employees are asked to rate various aspects of their work (e.g. Zwieten *et al.*, 2014; Kraan *et al.*, 2000), followed by department wide interventions (e.g. providing training).

As knowledge workers are relatively flexible in their work (when they do what and how they work), there is great potential for them to contribute to the improvement of their own well-being. New technologies are emerging, such as sensors available in smart phones, smart reasoning and e-coaching apps. In the SWELL project (Smart Reasoning for Well-being at Home and at Work (SWELL project 2015)) we see potential in using such new pervasive technologies to address well-being at work at an individual level (Koldijk, 2012). We see possibilities in using unobtrusive and easily available sensors to capture the knowledge workers behavior (e.g. computer interactions, webcam for facial expressions) and infer stress and the context in which it appears. Based upon this information, we aim to develop a system with a suite of support applications, that are context aware, i.e. optimally adapted to the situation and state of the user. Knowledge workers can then directly act, gaining a more healthy work style and preventing stress building up. Trends like ‘quantified self’ already show the

potential of collecting personal sensor data (e.g. heart rate, activity patterns) for health improvement. In their paper on technology for well-being, IJsselstein *et al.* (2006) describe that advancements in sensing and interpretation are promising. They further state that using technology for improving well-being has many advantages, e.g. its persistence or objectiveness, the possibility to provide just-in-time notifications with relevant, actionable information or their supportive and motivating role.

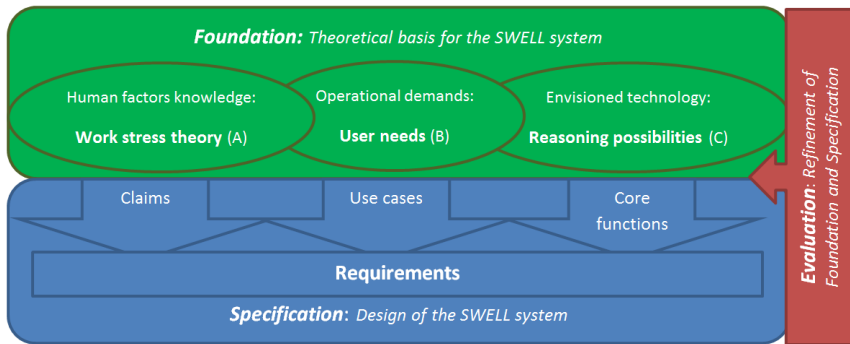


Figure 2.1: Situated Cognitive Engineering (sCE) approach.

To develop a theoretically and empirically grounded stress self-management system, we take a multi-disciplinary approach. By means of situated cognitive engineering (Neerinx and Lindenberg, 2008) we combine theory on work stress with input on user needs, taking in mind technological possibilities (see Figure 2.1). In this way, we generate a functional system specification with core functions and claims, which is then evaluated with users. The main focus of this chapter is the theoretical foundation. The general objective of the SWELL system is to improve well-being at work. An important question is: What defines well-being at work and what causes well-being? Many relevant theories are provided by several disciplines, e.g. Work Psychology, Biology or Behavioral Psychology. However, theories are diverse and different disciplines view the world from different angles, e.g. using different levels of abstraction. The big question is: How do different concepts relate to each other? One comprehensive and practical framework, that can be used as theoretical basis for the design of the envisioned self-management support, is still lacking. Moreover, psychological theories are often rather abstract and for implementing a solution many choices need to be made. We investigate the role of new technologies, which also provide new opportunities to study behavior, as well as new means to influence behavior.

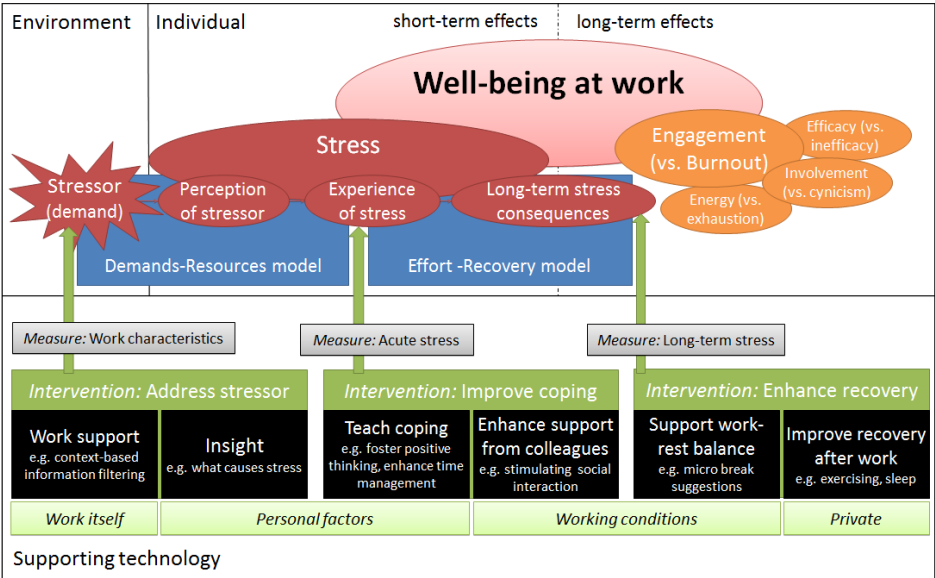


Figure 2.2: Our framework, combining various stress and intervention theories, as well as possibilities for real-time measurements and interventions with technology.

The main contribution of this chapter is therefore a general and pragmatic framework (see Figure 2.2), which combines various stress and intervention theories, as well as possibilities for real-time measurements and interventions with technology. This framework can be used for developing technologies addressing well-being at work, as is demonstrated in our SWELL use case. Moreover, we show that, vice versa, new technologies can also be used for theory building. Our research questions and the remainder of the chapter are structured around our framework in the following way:

- First, an initial study on user needs is shortly described, as a starting point for the system design.
- Then, we answer our *first research question*: Which concepts are relevant with respect to well-being at work? Several concepts related to well-being at work are presented: ‘burn-out’, ‘engagement’ and ‘stress’ (red/ orange parts).
- We then answer our *second research question*: Which person, work and context conditions can lead to negative stress? We present different work

stress models, that describe how stress in working environments is caused (blue parts).

- Then, we turn to our *third research question*: How can sensors be applied to automatically infer stress and the context in which it appears? We integrate knowledge on technical possibilities here (grey parts).
- Finally, we answer our *fourth research question*: Which interventions can be provided by means of pervasive technology to help a knowledge worker improve his well-being at work? Intervention theory for addressing work stress is presented (green parts). We also describe some technical possibilities for support (black parts).
- After having presented this framework, the envisioned system and first prototypes of technical support are presented, as well as results from evaluation studies with potential end users.
- We finish with a conclusions and discussion section, in which we present 7 research challenges that we identified.

## 2.2 Initial study on user needs

Following the situated cognitive engineering methodology (Neerinx and Lindenberg, 2008), we start with input from potential end users.

We held interviews with 5 knowledge workers who had experienced burn-out and organized a workshop with 7 employees to establish user needs. Knowledge workers indicated that the system should provide them an overview of performed work, preferably in combination with work behavior and the associated subjective experience. This information can then be used by the user to gain insight into work processes. For example, at the end of the day an overview could be provided on how time was spent and how stress evolved. Moreover, users indicated that they would want help in the form of tips. Ideally the tips are also well-timed, taking into account the user's current context. Finally, users indicated that the system could actively support them during their work. The system can take an active role in supporting the user, e.g. by filtering irrelevant emails or finding information relevant to the current work topic. We also identified some important factors to address, e.g. not irritating users and addressing privacy. Important is also the flexibility of the SWELL system, e.g. regarding the sensors used and the frequency of feedback.

This user input was used to guide the further design of the system. In the next sections we focus on important relevant domain knowledge.

### 2.3 Well-being at work concepts

In this section we aim to answer our *first research question*: Which concepts are relevant with respect to well-being at work? To answer this question, we performed a literature review (P. Vos, van Staalduinen, and Koldijk, 2010). The search engine ‘Web of science’ was used with the keywords: well being, commitment, satisfaction, stress and engagement. Based on 23 scientific publications an overview of the different concepts was made. The literature review revealed that there are many different related concepts and a many different models. Finally, the concepts ‘engagement’ and ‘stress’ were chosen, as they seemed most suitable to capture with sensors. In this section, we first describe the concept of engagement in more detail (see Figure 2.2). Then, literature regarding stress and its consequences is presented.

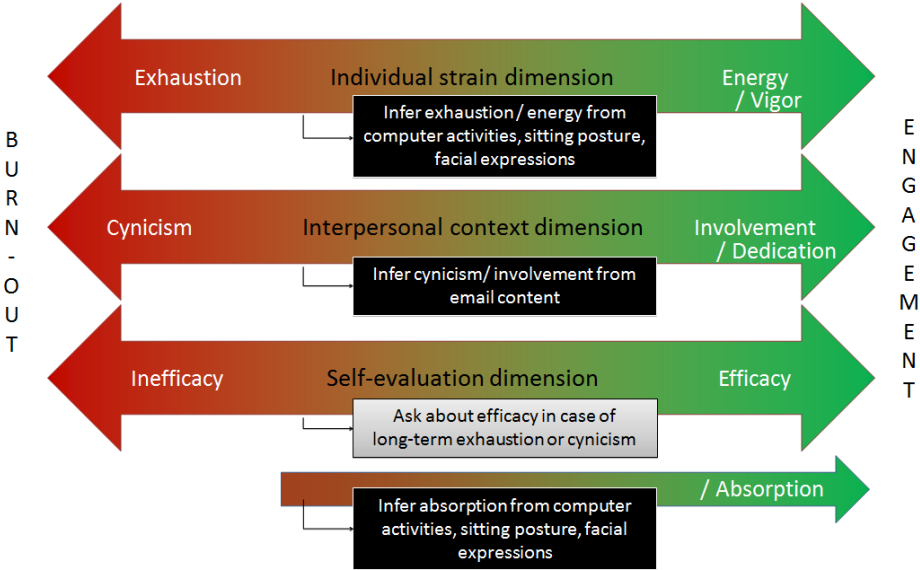


Figure 2.3: Well-being at work concepts ‘burnout’ and ‘engagement’, and ideas to infer certain aspects from captured (sensor) data.

#### 2.3.1 Engagement and burn-out

The relationship people have with their jobs can be described as a continuum between engagement and burn-out according to Maslach and Leiter (2008), see

Figure 2.3. They distinguish 3 dimensions: 1) Individual strain (exhaustion vs. energy), 2) Interpersonal context (cynicism vs. involvement), and 3) Self-evaluation (inefficacy vs. efficacy). According to this terminology, an engaged employee feels energy, involvement and efficacy. His state can be characterized as worrisome when he feels exhaustion (discouraged by chronic, overwhelming demands), cynicism (distance oneself emotionally and cognitively from one's work) and/or inefficacy (reduced sense of effectiveness/ personal accomplishment), which characterizes burnout (Maslach, W. B. Schaufeli, and Leiter, 2001). According to Maslach and Leiter (2008), a low score on burn-out means high engagement and vice versa. They further state that "engagement represents a desired goal for any burnout interventions." (p. 499).

W. B. Schaufeli *et al.* (2002) define burn-out and job engagement as two distinct concepts that can be measured independently. They describe engagement as the combination of vigor (high levels of energy and the willingness to invest effort), dedication (a sense of significance, enthusiasm and challenge) and absorption (being fully concentrated and deeply engrossed in one's work). The first two concepts are similar to those described by Maslach and Leiter (2008). The main difference lies in the third dimension, absorption, which is not the opposite of inefficacy, but a different aspect.

### 2.3.2 Stress

Besides engagement or burn-out, a relevant concept that can be experienced in the office from day to day is stress. In work by Le Fevre, Matheny, and Kolt (2003) we find that: "Selye (1964) was the first to use the term 'stress' to describe a set of physical and psychological responses to adverse conditions or influences." (p. 727). However, in research we find that the term stress is often used to refer to different things. Le Fevre, Matheny, and Kolt (2003) write:

"Selye (1956) used the term 'stressor' to describe the external force or influence acting on the individual and 'stress' to denote the resulting reaction, terminology adopted by many others (e.g. Code and Langan-Fox, 2001; Maslach, 1998; Quick et al., 2001). Some authors have used stress to denote such external forces and strain to denote the resulting reaction (e.g. Edwards, 1998), while others failed to clearly define how they were using the terminology at all (e.g. Smit and Schabracq, 1998; Wiholm et al., 2000). Further, some have simply used stress as a blanket term covering the whole process of external influence, appraisal, reaction, and effect (e.g. Deary et al., 1996)." (p. 728).



In our work we use the definition by Selye (see Figure 2.2). An environmental demand, or stressor, leads to a perception of the stressor, which is dependent on the particular characteristics of the individual. The individual’s perception of the stressor results in a particular experience of stress. An employee complaining about stress might thus mean that his working conditions are very demanding (the stressor), or that he feels that demands put upon him are higher than he can take (the perception of stressors) or that he feels stress reactions in his body (the experience of stress).

Selye (1975) distinguishes good stress (eustress) and bad stress (distress). Some amount of stress is not harmful and might even be beneficial to gain concentration and focus. Eustress occurs when the person experiences the right amount of demand. Distress occurs when a person experiences too much or too little demand. This is related to the Yerkes Dodson Law (Yerkes and Dodson, 1908), which describes that (empirically) performance improves with arousal, up to a certain point, after which it declines again.

Individual characteristics and appraisal play an important role in the experience of stress. The same stressor can be seen as problem, leading to negative emotions, causing distress, or as challenge, leading to positive emotions, causing eustress (Le Fevre, Matheny, and Kolt, 2003). This can depend on the amount of resources or feeling of control that the individual has. So even changing the mind-set of a knowledge worker could help him cope better with stressors. More details on the balance of demands and personal resources can be found in the section on work stress models.

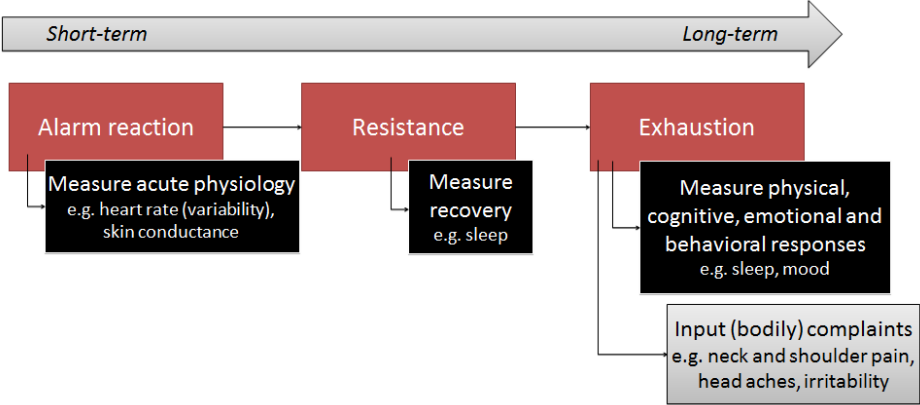


Figure 2.4: Stress reactions of the body, and measuring possibilities.

The body's short and long term reactions to stress can, from biological perspective, be captured in 3 stages (General Adaptation Syndrome; Selye, 1950, see Figure 2.4): 1) Alarm reaction - the fight or flight response; 2) Resistance - the body adapts to the stressor; and 3) Exhaustion - the body's resistance decreases due to long-term stress. The alarm reaction causes adrenaline to spread through the body and blood pressure rises (reaction of the nervous system). Under very stressful conditions, a shift in hormone production may take place, increasing stress hormones like cortisol, which increases blood sugar, but also suppresses the immune system (reaction of the hormonal system). This stress response system works well for dealing with short term stressors. When the stressor disappears the body gains back its natural balance. When the level of the stress hormone cortisol is high for a prolonged time, however, this has negative effects, e.g. on the brain. This shows the importance of recovery.

With lack of recovery, stress can accumulate and lead to health problems. Extended periods of stress can cause (see e.g. (Bressert, 2015)): Physical reactions (e.g. increased blood pressure, muscle tension, headache and sleeping problems); Cognitive reactions (e.g. problems with concentrating, problems with setting priorities and decreased efficiency in work); Emotions (e.g. irritation, feeling restless, tense and anxious); Changes in behavior (e.g. avoiding social contact, more risk taking, not being able to relax and increased complaining). Moreover, J. Bakker *et al.* (2012) explain that stress can not only directly lead to illness through its physiological effects, but also indirectly, through maladaptive health behavior, such as smoking, poor eating habits or lack of sleep.

Moreover, in Chronobiology, scientists describe ultradian and circadian bodily rhythms (e.g. Mejean *et al.*, 1988). Circadian rhythms describe the human's natural body rhythm throughout a day. Depending on daytime (light) and bodily processes (e.g. digestion), the human body traverses typical states with fluctuations from alert and active to relaxed and sleepy. There also seem to be morning types and evening types that differ in their states throughout the day. Ultradian rhythms describe recurrent cycles throughout the day. Rossi and Nimmons (1991) explain that every 90-120 minutes the body signals that it needs rest and change in activity. They describe that ignoring your body rhythm causes stress.

### 2.3.3 Conclusions: Relevant concepts for the system

In this section we aimed to answer our *first research question*: Which concepts are relevant with respect to well-being at work? We identified the following concepts (see Figure 2.2, orange/ red parts): engagement (vs. burn-out) and stress. We could assess the three underlying dimensions of engagement: energy, in-

vovement and efficacy or absorption. Especially energy (vs. exhaustion) might be an interesting concept to measure and monitor in real-time, as well as the dimension of absorption (e.g. a state like ‘flow’).

Moreover, we found that stress is a normal process and in form of eustress also good for well-being and performance. It cannot be the goal to prevent stress. Rather, employees should be helped to handle distress and prevent negative long term consequences. In SWELL the aim is to improve well-being at work, this means that we conjecture that a comprehensive approach that addresses all facets of stress is probably the most effective. A stressor might or might not lead to the perception of stress in a specific individual. So we could measure the stressor (e.g. work characteristics), as well as the individual’s perception of the stressor (e.g. acute stress). We could also take typical fluctuations over the day into account. In addition, we could analyze long term patterns in which stress is building up, which are even more dangerous. We could therefore measure recovery, e.g. sleep time or the amount of physical activity. Moreover, we could provide means to keep track of (bodily) complaints (e.g. headache, pain in the back or shoulders, experienced lack of motivation, irritability, restlessness etc.). Making employees aware of unhealthy stress patterns with lack of recovery may be important to prevent burn-out.

**Core functions of the system:** Based upon this part of the theoretical framework, we formulated the following core functions for the pervasive well-being system, together with the associated claim:

- F1: The SWELL system could collect information about: aspects of engagement, work characteristics, acute stress, and long-term stress/ recovery.
- Claim: This information is useful for data-driven and context-aware coaching.

## 2.4 Causes of work stress

After having described the concepts related to well-being at work, we now turn to models describing underlying causes. We aim to answer our *second research question*: Which person, work and context conditions can lead to negative stress? We present the four most influential work stress models, which all describe a balance between 2 variables, see Figure 2.5. The basic idea is that work becomes stressful when high demands are combined with: 1) insufficient resources (such as low job control and little social support), 2) little rewards, 3) little recovery,

or 4) an environment that mismatches with personal characteristics. We now outline each model in more detail. Based on each model, we identify aspects that can be addressed by means of technology. (Each identified technology has an identifier for later reference, see Table 2.1 for an overview.)

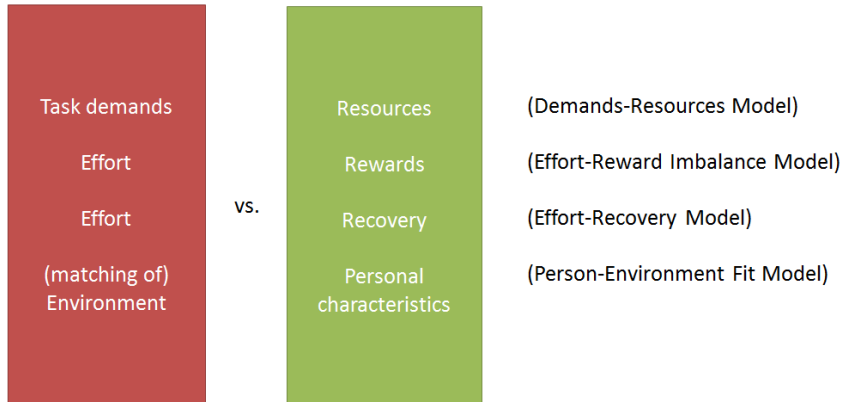


Figure 2.5: Different work stress models.

### 2.4.1 Job Demands-Resources Model

The first model can be characterized by a balance between job demands on the one hand and resources on the other hand (see Figure 2.6).

Karasek Jr (1979) developed the initial model called the Job Demands Control (JDC) model. This model states that, on the one side, the combination of high task demands with low control over the execution of tasks can lead to strain. On the other side, the combination of high demands with high control leads to learning and motivation. This means that high demands in itself are not bad, as long as the employee has enough control. Moreover, high demands can even be beneficial for learning and motivation, as long as job control is high.

The model was later later extended by Demerouti, A. B. Bakker, Nachreiner, *et al.* (2001) to the Job Demands-Resources model (JD-R model). Here the more general interplay between job demands and job resources is described (Demerouti, A. B. Bakker, Nachreiner, *et al.*, 2001):

- Job demands are aspects of the job that require effort. Examples are: physical workload, time pressure, emotional demands and the physical envi-

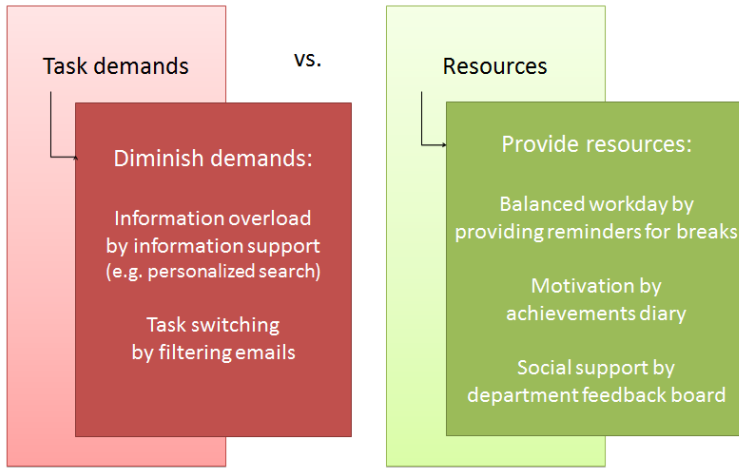


Figure 2.6: Job Demands-Resources model by Demerouti, A. B. Bakker, Nachreiner, *et al.* (2001), and possibilities for technological support.

ronment. High job demands in itself are not problematic; problems arise when the necessary resources lack.

- Job resources are aspects of the job that help in achieving work goals, reduce demands or stimulate personal growth and development. Examples are: autonomy, job control, social support (from colleagues, supervisor, family, peer groups), feedback, rewards, task variety and role clarity.

The model describes as well a process of health impairment as one of motivation (Demerouti and A. B. Bakker, 2011). Regarding health impairment, “demanding jobs or jobs with chronic job demands (e.g. work overload, emotional demands) exhaust employees’ mental and physical resources and may therefore lead to the depletion of energy (i.e. a state of exhaustion) and to health problems (e.g. general health and repetitive strain injury) (Bakker, Demerouti, & Schaufeli, 2003; Demerouti *et al.*, 2000, 2001; Leiter, 1993).” (p.2). So high job demands can lead to reduced health and energy. Having job resources can help employees to cope with high job demands. As Demerouti and A. B. Bakker (2011) state: “[job] resources may buffer the impact of job demands on job strain, including burnout (Bakker, Demerouti, & Euwema, 2005; Bakker *et al.*, 2003b; Xanthopoulou *et al.*, 2007b).” (p.2).

Besides being a buffer against health impairment, job resources also play a role in motivation: “Job resources may play an intrinsic motivational role be-

cause they foster employees' growth, learning and development, or they may play an extrinsic motivational role because they are instrumental in achieving work goals." (p.2). Also task demands play a role in motivation. As Demerouti and A. B. Bakker (2011) state: "There is a need for a challenge (i.e. a demanding condition) in order for job resources to be translated into task enjoyment and work engagement." (p. 3).

The WEB (Werkstressoren- Energiebronnen-Burnoutmodel) model (A. B. Bakker, W. Schaufeli, and Van Dierendonck, 2000) is another variant of the JD-R model, in which a direct link between demands, resources and the three aspects of burn-out is made: High demands cause exhaustion, whereas a lack of resources can lead to a decreased feeling of competence (inefficacy) and distancing oneself from work (cynicism).

**Supporting technology.** Based upon the Resources-Demands model, we can address the well-being at work from two sides: We can diminish the demands placed upon knowledge workers or provide additional resources.

Examples for diminishing demands: A typical demand on a knowledge worker is to deal with large amounts of information. We can make technology that can try to diminish information overload by providing information support, for example in the form of filtering context-relevant from irrelevant emails (Technology T01) or by enabling personalized search (T02). Another demanding aspect of the work is task switching. A computer tool could diminish this demand by helping employees to remain focused on the task at hand, e.g. by filtering irrelevant emails (T01 again) or with gamification, motivating employees to stay focused by giving points for less task switching (T03).

Examples for providing resources: A resource that the knowledge worker has is his motivation and self-efficacy. The computer tool can support motivation, e.g. by providing an achievements diary (T04), which is in line with work by Amabile and Kramer (2011) who showed that the feeling of making progress leads to more motivation and better performance. We could also facilitate social support, which is an important resource, e.g. facilitate support by peers by use of a department-wide feedback board (T05). Another resource is a good work-rest balance, with variation in tasks. The system could help to have a balanced workday by providing insights in what gives and costs energy, e.g. by providing an activity and workload overview, promoting better planning (T06). Taking enough recovery breaks could also be traced and supported with technology (T07). Important to take into consideration is keeping the knowledge worker in control and not posing additional demands.

### 2.4.2 Effort-Reward Imbalance Model

The Effort-Reward Imbalance (ERI) model by Siegrist (2012) can be characterized as a balance between effort on the one side and rewards on the other side. As long as the rewards are in balance with the efforts of the employee there is no problem. An imbalance might occur when the employee's efforts are higher than his rewards, which might happen for example due to overcommitment. Such an imbalance may result in stress and negative consequences for health.

**Supporting technology.** Based upon the Effort-Reward Imbalance model, we can address well-being at work by helping employees to match their efforts to the expected rewards. We might for example support realistic goal setting and in this way diminish pressure and disappointments. Insight regarding planned time versus the real time may facilitate better (re)planning and setting more realistic goals (T06 again). Moreover, looking back at ones achievements could help employees to get a better feeling of their productivity (T04 again). Also aspects of gamification might provide employees small motivating rewards, e.g. collecting points for staying focused (T03 again).

### 2.4.3 Effort-Recovery Model

The Effort-Recovery (E-R) model by Meijman *et al.* (1998) can be characterized as a balance between effort and recovery, see Figure 2.7.

In their model they describe that job demands and resources lead to negative strain during work. After work, home demands and resources lead to strain reactions when home. The individual can perform activities which can have a positive effect on recovery, leading to a particular psychological and energetic state at bedtime. By means of sleep, additional recovery can be gained and the individual starts the next workday with a certain psychological and energetic state before work. Failing to recover enough from strain can make the experience of work demands the next day higher and the experienced resources lower, leading to even more strain. This process can be a vicious circle. According to Demerouti, A. B. Bakker, Geurts, *et al.* (2009) lack of recovery can "result in an accumulative process developing into chronic load reactions (or 'allostatic load' according to McEwen's (1998) allostatic load theory), such as chronically elevated heart rate, hypertension, chronic fatigue, and persistent sleep problems (Sluiter, Frings-Dresen, Van der Beek, & Meijman, 2001)." (p. 88).

Four important dimensions play a role in recovery (Sonnentag and Fritz, 2007): psychological detachment, relaxation, mastery and control. Psychological detachment from work can bring the psychophysical system back to its normal

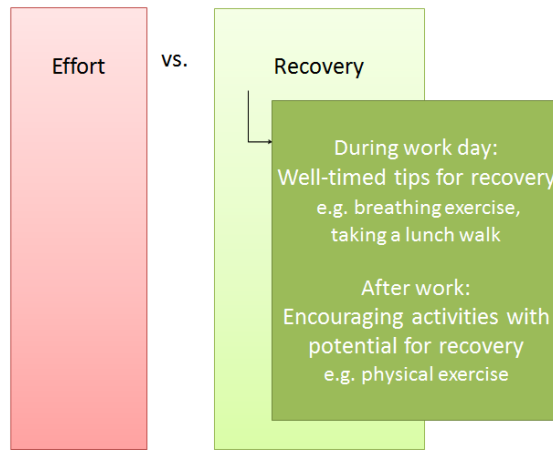


Figure 2.7: Effort-Recovery model by Meijman *et al.* (1998), and possibilities for technological support.

state, improving recovery. Relaxation causes decrease in physical activation and more positive affect, also improving recovery. Controlling what activity to perform can improve esteem and efficacy, which might have a positive effect on recovery. Mastery in performing challenging activities can cause improvement of skills, competence and esteem, which also might have a positive effect on recovery. Demerouti, A. B. Bakker, Geurts, *et al.* (2009) describe that “it seems important that people engage in activities that appeal to other systems than already used during work, and that are not (again) stressful.” (p. 88). In Demerouti, A. B. Bakker, Geurts, *et al.* (2009) we further find: “Recovery experience refers to the degree to which the individual perceives that the activities he/she pursues during non-work time helps him/her to restore energy resources (Sonnentag & Natter, 2004).” (p. 91). And further: “In terms of physiological indicators, it means that blood pressure and heart rate are reduced in the evening (Rau, 2006). In terms of psychological indicators, the recovery process during non-work time has to do with less rumination (Cropley *et al.*, 2006), or better well-being before going to bed (Sonnentag, 2001)” (p. 112).

In general, physical activity seems to be a good means for recovery. Research by Norris, D. Carroll, and Cochrane (1992) showed that “in an adolescent population aerobic training does appear to provide some benefits with regard to psychological stress and well-being” (p.64). Hassmen, Koivula, and Uutela (2000) found that “individuals who exercised at least two to three times a week experi-



enced significantly less depression, anger, cynical distrust, and stress than those exercising less frequently or not at all.” (p. 17). Moreover, Demerouti, A. B. Bakker, Geurts, *et al.* (2009) present an overview of research findings regarding recovery indicators and their effects, for example: psychological detachment during social or physical activities has a positive effect on mood; time spent on physical activities improves vigor; sleep quality decreases fatigue; work related activities cause a decrease in well-being.

**Supporting technology.** Based upon the Effort-Recovery model, we can address well-being at work by making employees aware that recovery during work and non-work time is very important. Interventions could be aimed at taking well-timed breaks during the work day (again T07). Passive, as well as active breaks could be suggested, e.g. relaxation or taking a lunch walk. On the other side, an important aspect of improving well-being at work is also what someone does in his free time. We see that activities after work give potential for recovery. This model is interesting within the SWELL project, as it can combine the domains of well-being at work and at home. Interventions for more well-being could be aimed at better relaxation or detaching from work, e.g. by means of a hobby (T08). Addressing physical fitness could also be a good intervention (T09).

#### 2.4.4 Person-Environment Fit Model

The Person-Environment (P-E) Fit model was initially proposed by French, Rodgers, and Cobb (1974) and describes a fit between person and environment characteristics. A misfit between the person and his environment can lead to strain, with the danger of illness. There can for example be a misfit between personal abilities and environmental demands or between personal needs and environmental supplies (Caplan, 1987). Leiter and Maslach (2003) developed the Areas of Worklife Scale (AWS) around this idea. They say that “the greater the perceived gap between the person and the job, the greater the likelihood of burnout; conversely, the greater the consistency, the greater the likelihood of engagement with work” (p. 101). The AWS has items on 6 aspects: workload, control, reward, community, fairness and values.

**Supporting technology.** Based upon the Person-Environment Fit model, we can address well-being at work by helping employees realize that performing tasks that fit their personal preference is very important for their well-being. Tasks that give energy and tasks that cost energy could be identified by provid-

ing an overview over tasks and energy levels over the day (again T06). In future, the employee can then try to find work fitting his preferences more.

#### 2.4.5 Conclusions: Addressing causes of work stress

In this section we aimed to answer our *second research question*: Which personal, work and context conditions can lead to negative stress? We elaborated on several work stress models, that describe how stress in working environments is caused. The different models all have a different focus and complement each other. There are no specific personal, work or context conditions that generally lead to stress. Work becomes stressful when high demands are combined with: 1) insufficient resources, 2) little rewards, 3) little recovery, or 4) an environment that mismatches with personal characteristics. The most useful models for developing pervasive systems are the Job Resources-Demands model and the Effort-Recovery model, which we integrated into our framework (see Figure 2.2, blue parts). The Job Resources-Demands model describes how (environmental) stressors can cause the experience of stress. The Effort-Recovery model describes how the experience of stress can lead to long term stress consequences. We presented several ideas how technology can diminish demands, enhance resources or help with recovery. Table 2.1 provides an overview of identified technologies, the underlying models, and the associated claims.

Note that all models describe work stress in qualitative terms. Our aim is to quantify several aspects by using sensors. Demands could be quantified by measuring work characteristics (e.g. tasks and content worked on). Personal resources could be quantified by measuring the associated acute stress (e.g. physiological stress responses, mental effort). Recovery of the individual could be quantified by measuring long-term stress aspects (e.g. sleep time, physical activity). We elaborate on inferring these aspects from sensor data in the next section.

### 2.5 Inferring stress and its context

After having described concepts related to well-being at work, and causes of work stress, we now focus on assessing stress and its context. In current practices, most often questionnaires are being used (e.g. Zwieten *et al.*, 2014; Kraan *et al.*, 2000). However, this data collection has several shortcomings: Data is self-reported, suffering from recall bias and subjectivity, and data is only collected once in a year for example. Using sensors overcomes these shortcomings. Real-time measurements in the office provide a more objective measurement of stress-related variables. Moreover, collecting data about stress together with the

Table 2.1: Overview of identified technologies and associated claims.

ID	Possibility for technological support	Underlying theory	Claim
T01	Filtering emails	JD-R Model	Diminishes demands by reducing information overload.
T02	Personalized search	JD-R Model	Diminishes demands by reducing information overload.
T03	Gamification facilitating focus	JD-R Model, ERI model	Diminishes demands by diminishing fragmentation, enhances motivation by means of small rewards.
T04	Achievements diary	JD-R Model, ERI Model	Enhances resources or rewards by fostering motivation.
T05	Department-wide feedback board for peer support	JD-R Model	Enhances resources by means of social support.
T06	Activity and workload overview for insight	JD-R Model, ERI Model, P-E Fit Model	Provides insight in the balance between demands/ resources, efforts/ rewards or person-environment fit.
T07	E-coach for taking enough recovery breaks	JD-R Model, E-R Model	Enhances resources or recovery by taking rest breaks.
T08	E-coach for relaxation or detaching after work	E-R Model	Enhances recovery by detaching.
T09	E-coach addressing physical fitness	E-R Model	Enhances recovery by releasing stress with physical activity.

context in which it appears can give insights that can more directly be acted upon by an employee. This gives much richer information than annual questionnaires.

Therefore, we now aim to answer our *third research question*: How can sensors be applied to automatically infer stress and the context in which it appears? We focus on (physically) unobtrusive, relatively cheap sensors that can easily be used in office environments. Following the situated cognitive engineering methodology (Neerincx and Lindenberg, 2008), we integrate knowledge on technical possibilities here. We also investigate user choices regarding data collection.

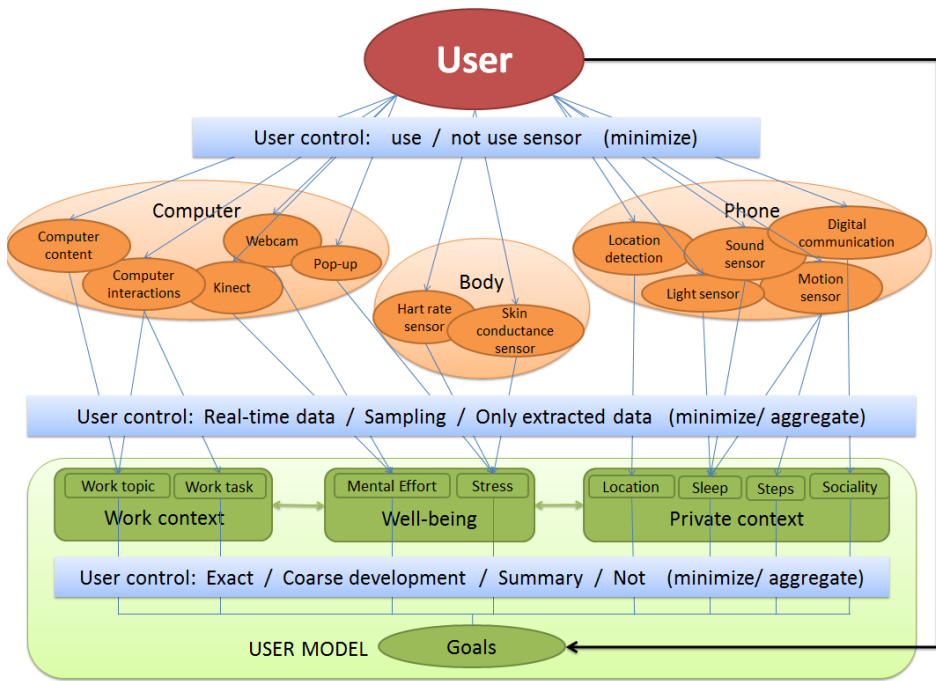


Figure 2.8: Overview of the system and its user model, which holds information on the users work context and his well-being. Moreover, information on the private context may be included. The user can decide on not using particular sensors and restrict in which detail data is collected.

### 2.5.1 Technical possibilities

In the previous sections, we identified several relevant concepts that the system could measure to provide data-driven coaching and context-aware support: work characteristics, acute stress, long-term stress/ recovery, and aspects of engagement. In Figure 2.8 we present an overview of the types of information and the sensors that can be used in the pervasive system to infer these aspects.

**Work characteristics.** First of all, we can measure work characteristics. The task (e.g. write report, search information) someone is performing can be inferred from computer interaction data. We present algorithms for real time task inference in Koldijk, van Staaldin, Neerincx, *et al.* (2012)<sup>1</sup>. Moreover, which project someone is working on can be detected by analyzing the content of accessed documents and websites. Algorithms for topic detection are presented in Sappelli (2016). The combination of tasks and topics can provide valuable information on the context in which stress appears. Based upon information on what someone was working on when, we can also infer the amount of task switching, variation in tasks, and the work-rest-balance. Most informative are probably deviations from usual behavior of the specific user.

**Acute stress** With respect to inferring of stress from sensor data, Sharma and Gedeon (2012) provide a compact survey. Often, body sensors are used to measure the physiological stress response directly, e.g. skin conductance (J. Bakker *et al.*, 2012) or heart rate (Hogervorst, Brouwer, and W. K. Vos, 2013). More and more unobtrusive devices are entering the market, like measuring watches, so this might be a potentially interesting measure to use. As a critical side note however, these devices may not be accurate enough to determine the more insightful variable of heart rate variability (HRV). Moreover, many external influences on physiology exists, e.g. drinking coffee or physical activity. Asking the user himself for input on stress may be useful.

There also is potential in using outward characteristics, such as facial expressions, postures or computer interactions as indicators for the user's mental state. Facial expressions are currently mainly used for inferring emotions, but facial expressions could also show cues to infer other mental states that might be more relevant in a working context. In earlier work, where working conditions were manipulated with stressors, we found that specific facial action units may be indicative of experienced mental effort (Koldijk, Sappelli, Neerincx, *et al.*, 2013). Research by Dinges *et al.* (2005) suggest that facial activity in mouth and eye-brow regions could be used to detect stress. Moreover, Craig *et al.* (2008) looked

---

<sup>1</sup>Chapter 4

at facial expressions while students worked with an online tutoring system. Association rule mining identified that frustration and confusion were associated with specific facial activity. Mental states are also being estimated from computer interaction data. Results by Vizer, Zhou, and Sears (2009) indicate that stress can produce changes in typing patterns. Finally, Kapoor and Picard (2005) describe work on recognizing interest in students by means of computer interaction and posture. Currently, we are therefore also investigating in how far we can infer stress or experienced mental effort from facial expressions, computer interactions and postures (Koldijk, Sappelli, Neerincx, *et al.*, 2013). Due to individual differences, general models will have to be adapted to the specific user for reliable estimates.

**Long-term stress/ recovery** To measure the more long-term physical, cognitive, emotional and behavioral responses, as well as recovery from stress (see Figure 2.4), it may be interesting to include private aspects. With mobile phone sensors, a rough estimate of sleep time can be provided by the combination of darkness, silence and recharging of the phone battery (see e.g. (Goalie 2015)). Moreover, the amount of physical exercise, which is a good relief for stress, can be measured by means of sensors (e.g. via mobile phone (Goalie 2015), via band (Fitbit 2015)). Moreover, a very rough estimate of sociality can be made, based upon the amount of phone communication. Besides, location information (e.g. GPS) can be useful, e.g. enhance the timing of feedback.

**Aspects of engagement** Besides the aspects already included in Figure 2.8, we have some initial ideas to measure certain aspects of engagement (see Figure 2.3) during work. Based on sensor data, energy (vs. exhaustion) may be a concept that can be inferred, e.g. by looking at someones sitting posture, computer interactions, or maybe facial expressions. This could give longitudinal information on the individual strain of an employee. Moreover, we could get a first indication of involvement (vs. cynicism) from textual analysis of email content. The concept of efficacy (vs. inefficacy) however, might probably best be assessed with questions to the knowledge worker. For example, when the longitudinal data shows little energy, the employee might want to fill in some questions on feelings about his efficacy, to be able to give an early warning and provide help in time. Finally, a state of absorption, like ‘flow’, might be detectable based on sensor data. A high focus on the task at hand, might be recognizable based on computer behavior (e.g. focus on one application), typical postures (e.g. leaning forward, sitting still) or facial expressions.

### 2.5.2 User choices regarding data collection

To estimate the identified states, various sensors are necessary, see Figure 2.8. Applying sensor technology to monitor personal activities most probably raises concerns related to privacy. Therefore, we performed a user study to investigate what the general perception of using various types of information and sensors is.<sup>2</sup> Nine participants tested a sensing and e-coaching prototype for two weeks. In a questionnaire, they were then asked to set the configurations for data collection to be used for own insight and for improving the e-coaching app.

We found that some sensors are in general perceived as more privacy sensitive (e.g. webcam, sound sensor, computer content, digital communication), others as less privacy sensitive (e.g. motion sensors, heart rate, skin conductance). However, preferences regarding data collection are diverse and depend on the goal for which they want to use the system and the trade-offs they make for themselves regarding privacy. The system should therefore be configurable, such that the user can 1) decide which sensors to use, 2) decide in which detail information is extracted from the sensors, and 3) decide to store information in exact or only aggregated form (see Figure 2.8). Users may want to experiment how much functionality they can gain with disclosing certain types of data.

### 2.5.3 Conclusions: Using sensing and reasoning

In this section we aimed to answer our *third research question*: How can sensors be applied to automatically infer stress and the context in which it appears? We provide an overview of all possibilities for real-time measurements in Table 2.2. The user study showed that user's are only interested to collect data that is necessary for supporting their specific goal, so the system should be configurable.

**Core functions of the system:** We now sum up the identified core functions of the system, together with the associated claims:

- F1.1: The SWELL system shall infer relevant information from unobtrusive sensors to provide real-time objective measurements.
- Claim: Sensors provide real-time information on stress and the context in which it appears, which the employee can directly act upon.
- F1.2: The SWELL system shall only collect data that is necessary to support the user's goal.

---

<sup>2</sup>We briefly describe the main findings here, for more details we refer to Chapter 7.

- Claim: User's are only willing to collect information relevant to their personal goal (due to privacy).

Table 2.2: Overview. From left to right: The 3 aspects in the stress chain. For each aspect, several indicative factors can be measured (upper part), and different (technology based) interventions can be provided (lower part).

Problem	Stressors: 'My environment poses high demands.'	Experience of stress: 'I feel I cannot handle all demands.'	Stress consequences: 'I experience stress symptoms.'
Measure	Work characteristics	Acute stress	Long-term stress/ recovery
Concept & how to infer	<b>Tasks and content worked on:</b> computer activity. <b>Variation in tasks, task switching, work-rest balance:</b> computer activity (also calendar).	<b>Physiological stress responses:</b> skin conductance and heart rate (variability) from measuring watch. <b>Mental effort:</b> infer from facial expressions, posture, computer activity.	<b>Sleep time:</b> mobile phone sensing, e.g. using the combination of silence, darkness and recharging of the phone battery. <b>Physical activity</b> accelerometer, GPS.
Intervention	Address stressors (primary prevention)	Enhance coping (secondary prevention)	Enhance recovery (tertiary prevention)
Example technology	<b>Providing work support:</b> e.g. filtering emails (T01) and personalized search (T02). <b>Providing insight in the sources of stress:</b> e.g. activity and workload overview (T06).	<b>Helping to improve coping abilities:</b> e.g. gamification for focus (T03) and achievements diary (T04). <b>Fostering support by colleagues:</b> e.g. department-wide feedback for peer support (T05).	<b>Supporting work-rest balance:</b> e.g. e-coach for recovery breaks (T07). <b>Helping to improve recovery after work:</b> e.g. e-coach for detaching after work (T08) and e-coach for physical fitness (T09).

## 2.6 Improving well-being at work

We now have described concepts related to well-being at work, causes that play a role in the experience of stress, and means to assess relevant aspects with sensors. As a next step we aim to find an answers to our *fourth research question*: Which interventions can be provided by means of pervasive technology to help a knowledge worker improve his well-being at work? We describe intervention and behavioral change theory.



### 2.6.1 Intervention theory

There are different possibilities to address well-being at work and diminish stress. First of all, one can distinguish prevention approaches aimed at *different stages* in the stress chain (Ivancevich *et al.* (1990); see Figure 2.2, upper green parts). Primary prevention is aimed at the stressors, e.g. changing the work or work situation, to prevent risks. Secondary prevention is aimed at the (short-term) stress reactions, e.g. helping employees to develop good coping strategies, to handle stress risks and their consequences. Tertiary prevention is aimed at addressing (long-term) stress consequences, e.g. promoting a balanced life style, to recover.

Moreover, interventions can target *different areas* (see Figure 2.2, lower green parts). Based on literature, we identified four areas: the work itself, personal factors, the working conditions and private circumstances (P. Vos, van Staalduinen, and Koldijk, 2010). To support the employee to reach more well-being, the intervention should be targeted at the problem area. One could, first of all, change the work itself, improve work planning, or get a more focused workflow. Secondly, the intervention can target personal factors. One could enhance self-knowledge (e.g. what causes my stress), or improve active coping. Fourth, the intervention can target working conditions. One can address organizational aspects, social aspects (e.g. support from colleagues), or the work-rest balance. Finally, the intervention can address private circumstances. One can address social aspects (e.g. support from friends), or recovery.

Finally, we can distinguish *various types* of stress reducing interventions (e.g. Richardson and Rothstein, 2008). The most suitable type of intervention may depend on the employee's preference: cognitive-behavioral (e.g. coping skills and being more assertive), creativity, exercise, food, journaling, relaxation, social, or time-management/organizational. Note that an intervention can e.g. be social and creative at the same time.

### 2.6.2 Behavioral change

Until now, we explained what aspects interventions may address to improve well-being at work. However, changing the behavior of an individual may be difficult, especially in case of (bad) habits. Therefore we now consider behavioral change theory (see e.g. (Peters, 2015)).

People may know that particular behavior may be good for them, but still they may sustain their old behavior. Fogg (2002) identified three main hurdles preventing humans to perform the right or healthy behavior: lack of ability, lack of motivation and lack of a well-timed trigger. The interventions should

be designed in a way that they address these hurdles. More specific relevant determinants to address are:

- Regarding the *motivation* to perform the desired behavior:
  - Risk awareness. An employee could behave in a certain manner because he is not aware of the risks of his current behavior. The system could make employees more aware of health risks in order to initiate behavioral change.
  - Motivation. The employee might be not motivated enough to change something about his behavior. The system could motivate the employee, e.g. by coupling good behavior to collecting points (gamification).
  - Social influences. The behavior of a particular individual is also influenced by his social surrounding. Particular norms might be in place. The system could help employees to realize what realistic norms are and motivate them to listen to their own body.
- Regarding the *ability* to perform the desired behavior:
  - Skills. The employee might lack important skills. The system could help the employee by providing suggestions (e.g. ideas for short breaks).
  - Supportive environment. A supportive environment has a positive effect on behavioral change. The system could be supportive e.g. by filtering emails to prevent distractions.
  - Self-efficacy. The employee might have the feeling that he is not up to the task at hand. The system could improve self-efficacy e.g. in form of positive, motivating feedback.
- Regarding a well-timed *trigger* to perform the desired behavior:
  - Attention. The individual might just lack attention to perform the behavior at the appropriate occasion. The system could provide well-timed reminders to action.
  - Behavioral awareness. The employee might not be aware of all his behavior, e.g. he might not be aware of how often he checks his mail or how infrequently he takes breaks. The system could make the employee more aware of his behavior by mirroring his activities to him (feedback).

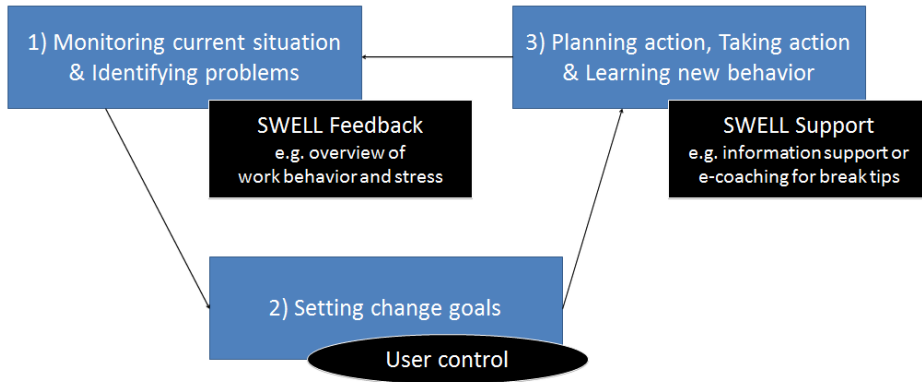


Figure 2.9: Behavior change and how technology could support.

For someone to successfully change his behavior, the following 3 main aspects should be supported in the system (see Figure 2.9): 1) Monitoring current situation and identifying problems; 2) Setting change goals and planning action; and 3) Taking action and learning new behavior. First, the system supports the knowledge worker in finding areas to address. As starting point the user can do an assessment. Furthermore, the system will provide feedback and insights into current behavior. Then, the system provides change goals, depending on the specific problem that was identified (e.g. diminish stressors, change personal coping, improve recovery). Finally, the system helps the user in reaching his goals by providing (work) support or recommending specific actions to perform. Recommendations will be personalized, which means that the system learns which kind of recommendations are preferred. This process (1-3) is iterative.

Moreover, we identified the most appropriate Behavior Change Techniques (Michie *et al.*, 2008) for the pervasive system, based on the list presented in Korte *et al.* (2014). These are: feedback, self-monitoring, contextual risk communication, and reminders or cues to action.

### 2.6.3 Conclusions: Technology based interventions

In this section we aimed to answer our *fourth research question*: Which interventions can be provided by means of pervasive technology to help a knowledge worker improve his well-being at work? The system can address different stages in the stress chain (see Figure 2.2): the stressor, improving coping and enhancing

recovery. We also found several areas to address: the work itself, personal factors, working conditions or private aspects. Based upon this categorization we added some general ideas for technology based interventions to our framework (Figure 2.2, black parts): An intervention focused at the stressor and the work itself, is e.g. providing work support. An intervention focused at the stressor and personal factors, is e.g. providing the knowledge worker insight in the sources of his stress. Furthermore, an intervention focused at coping and personal factors is e.g. helping the employee to improve his coping abilities. An intervention focused at coping and the working conditions, is e.g. fostering the support by colleagues. Finally, an intervention focused at recovery and working conditions, is e.g. supporting a work-rest balance throughout the workday. An intervention focused at recovery and private circumstances, is e.g. helping the employee to improve recovery after work. Based on the general theoretical framework, also further technology supported interventions can be designed. In Figure 2.10 we show how the specific supporting technologies (T) identified in the section on work stress models can be placed into this framework.

In designing a pervasive system to address well-being at work, simply implementing relevant interventions might not be sufficient. The system should support the employee throughout the behavioral change chain: it should help with monitoring the current situation and identifying problems. The user should be able to set change goals and determine the plan of action supported by the system. The system then can help the user to take action and learn new behavior. Also barriers towards changing behavior should be taken in mind. With pervasive technology we could improve the employee's motivation to change behavior, enhance the ability of knowledge workers to perform the desired behavior, or provide triggers to initiate action.

**Core functions of the system:** We now sum up the identified core functions of the system, based upon this part of the theoretical framework, together with the associated claims:

- F2: The SWELL system shall address 3 different causes of stress: address the stressor (F2.1), coping (F2.2) and recovery (F2.3).
- Claim: By providing different types of interventions, different causes of stress can be addressed with the system, making it usable in more situations.
- F3: The SWELL system shall foster behavioral change by: helping to monitor the current situation and identifying

problems (F3.1), letting the user set personal goals and enable specific functionality (F3.2), and helping to learn new behavior, by fostering the ability, motivation or trigger to take action (F3.3).

- Claim: By using behavior change theory the system will be more effective in actually bringing about behavioral change regarding well-being at work.

## 2.7 Envisioned system and evaluation of prototypes

Throughout this chapter, we formulated several core functions for the system. We sum them up here. The envisioned pervasive SWELL system supports the knowledge worker to improve well-being at work (OBJ). The SWELL system could collect information about: aspects of engagement, work characteristics, acute stress, and long-term stress/ recovery (F1). The SWELL system shall infer relevant information from unobtrusive sensors to provide real-time objective measurements (F1.1). The system only collects data that is necessary to support the user's goal (F1.2). With respect to behavioral change, the user will start with getting insight in his situation and identifying problems that he wants to address (F3.1). Based on these insights the user can then set personal goals and enable specific desired SWELL functionality (F3.2). In case the environment poses high demands, the user may decide to address some of his stressors (F2.1). In case the user feels overwhelmed by demands placed upon him, he may decide to address some of his coping abilities (F2.2). In case the employee experiences stress symptoms, he may decide to enhance recovery (F2.3). Behavior change techniques are used to foster motivation, ability and triggers to take action (F3.3).

In the next sections we describe first prototypes of different SWELL functionality. Figure 2.10 shows how the prototypes fall into our framework. All systems are aimed at improving well-being at work. Most prototypes make use of sensor information.

- The SWELL Workload Mirror tries to tackle stress in the beginning of the stress chain (e.g. 'what causes stress?') with the aim of helping employees to address the stressor itself. It is an implementation of T06 "activity and workload overview" and provides insights regarding stress and the context in which it appears. Based on these insights, the user might want to use one of the other SWELL systems for support.
- The SWELL HappyWorker system tries to tackle stress in the beginning of the stress chain (diminishing demands) with the aim of addressing the

stressor itself. It is an implementation of T02 “personalized search” and helps employees find relevant information.

- The SWELL Fishualization tries to tackle stress in the middle of the stress chain, helping employees to cope with stress. It is an implementation of T05 “department-wide feedback for peer support” and is aimed at fostering awareness and communication about stress at work.
- The SWELL NiceWork app tries to tackle stress in the middle and end of the stress chain. It is an implementation of T07 “e-coach for recovery breaks” and provides interventions aimed at improving coping, and enhancing recovery.

We now describe the prototypes in more detail.<sup>3</sup> We only briefly sum up the main functionality here. Our contribution is describing the prototypes in terms of our general framework. We present feedback that we got on our prototypes by potential end users in small-scale user studies.

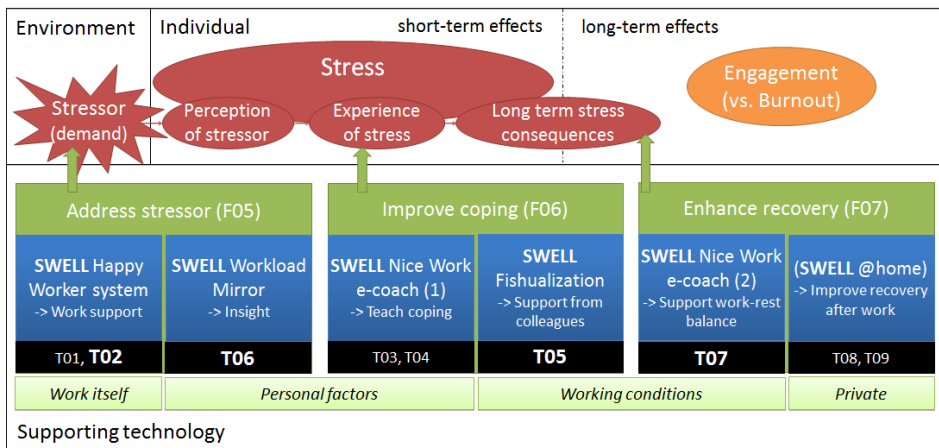


Figure 2.10: SWELL system functionality in our general framework.

### 2.7.1 Insight in stress sources - SWELL Workload Mirror

We here shortly present the SWELL WorkloadMirror. (For details we refer to the original work presented in Koldijk, Koot, *et al.* (2014)).

<sup>3</sup>I mainly worked on the SWELL Workload Mirror, the SWELL Fishualization was a cooperation with some colleagues, the SWELL NiceWork app was work of my intern, and the SWELL HappyWorker system was developed by colleagues.

The SWELL WorkloadMirror provides insights in work behavior and well-being (T06) (see Figure 2.11). How much time did the worker spend on which activities, related to which projects? During which time slots did the worker experience stress? By depicting objective measurement data, the user can relate his stress level to causes. By self-reflection, the employee could become aware of long term patterns, e.g. stress building up due to lack of recovery. He could formulate specific well-being goals to work on with an e-coach.

Regarding our framework, the SWELL WorkloadMirror, which provides insight, is aimed at personal factors. It tries to tackle stress in the beginning of the stress chain (e.g. ‘what causes stress?’) with the aim of helping employees to address the stressor itself (primary prevention). Its main basis is the Job Resources-Demands model; the employee can get insight in what demands cause stress and the importance of resources. It measure work characteristics and acute stress. With respect to behavioral change it helps with monitoring the current situation and identifying problems. The behavioral change techniques ‘feedback’ and ‘self-monitoring’ can increase behavioral awareness and therefore provide a trigger to take action.

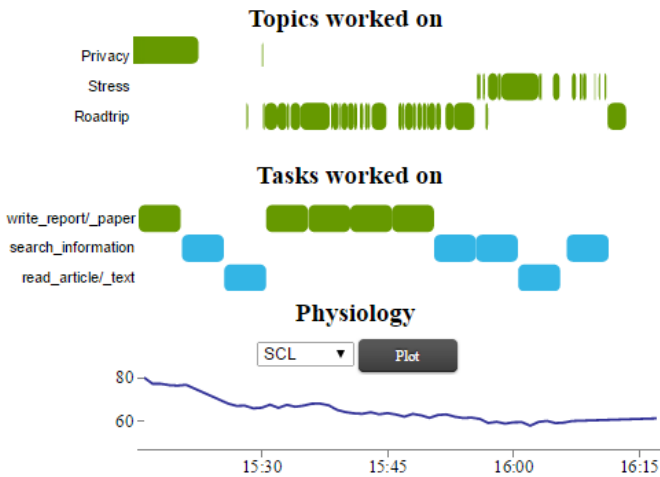


Figure 2.11: SWELL WorkloadMirror. Initial idea on visualizations for feedback.

**Feedback on our ideas.** A preliminary qualitative user study in form of a workshop was held with 18 young professionals from the The Hague region who visited TNO. First, the theoretical framework and the general idea of the prototypes were presented to these potential end users. Then, the group was

split and each subgroup was asked to give feedback on one of the prototypes. We asked them to first write down their own feedback and then discuss it with the group, to make a general top 3 of aspects that are top, and a top 3 of aspects that they would change or add.

6 participants (2 female, average age = 33.7) gave feedback on the SWELL Workload Mirror. The aspects that the group liked most were: 1) The system not only gives insight in stress, but also in how you work. 2) The system provides insight which is useful for business planning. 3) You can test yourself: “which effect does it have when I act like this”. Moreover they said that the prototype helps to focus on what gives energy, working more motivated and happy. The system would be preventive (“get insight quickly, otherwise you’re too late”). Finally, models of human behavior could be made based on the data, for research.

When asked what they would want to change or add, the main aspects mentioned were: 1) Getting personalized tips (feedback, suggestions). 2) Keeping it simple (non-experts may misinterpret the data) or personalizing the visualizations (simple vs. complicated). 3) Using gamification and rewards. Moreover, they said that the prototype only causes changes in awareness, not yet in other aspects of behavioral change. The focus should be on improving, instead of feedback on how bad the current situation is. Furthermore, the system currently lacks causal relations between tasks and stress. They also said that it would be interesting to add benchmarking with colleagues. Finally, participants expressed concerns regarding privacy. (Regarding the manner in which feedback is given, Hartsuiker (2015) further investigated feedback with avatars instead of graphs.)

**Evaluation study.** As the SWELL WorkloadMirror may be perceived as rather obtrusive (collecting all kinds of sensor data), we decided to investigate hurdles to use the system.<sup>4</sup> In a session with 11 knowledge workers who took part in a department meeting at TNO, first a presentation of a scenario for the SWELL Workload Mirror was given. Then, the participants were asked to write down all their potential concerns with the system.

In summary the hurdles were the following. Many participants had concerns about who could access their data. Many explicitly mentioned that they would not want the data to be shared with the management. Many participants were also afraid that the system would require effort, which might not outweigh its benefits. About half of the participants mentioned that they would want to know exactly what happens with the data, e.g. what is stored and where. About half of the participants had concerns regarding the performance of the system, e.g.

---

<sup>4</sup>We briefly describe the main findings here, for more details we refer to Chapter 7.



slowing down the computer. About a third of the participants had doubts about the system, in the sense that they were not sure whether it would solve their problem. About a fourth of the participants had issues regarding the quality, e.g. the reliability of the inferences made by the system. About a fourth mentioned issues regarding their control over the system. Finally, some participants mentioned that they would feel monitored.

Concerns regarding the effort the system would require highlight the important of automatic inferences and smart support, while users also wish to stay in control. A human-in-the-loop approach, combining automatic processing with human interaction seems a good solution. Furthermore, inferences should be reliable, while not slowing the computer down. Solutions can be using simple algorithms, running inference algorithms on a server or just analyzing samples of data. Some users express doubts about the SWELL system, so this solution to more well-being at work may not be suitable for every user. Moreover, we found that many concerns are related to privacy, i.e. the issues of: who can see the data, what will happen with the data, sharing data with the management and the feeling of being watched. As a result, we addressed these privacy concerns in more detail. We performed a privacy impact assessment and investigated how we can apply privacy by design. (For more information on integrating privacy aspects in system design see (Koldijk, Koot, *et al.*, 2014)<sup>5</sup>.)

### 2.7.2 Fostering colleague support - SWELL Fishualization

We here shortly present the SWELL Fishualization. (For details we refer to the original work presented in Schavemaker, Boertjes, *et al.* (2014)<sup>6</sup> and Schavemaker, Koldijk, and Boertjes (2014)).

The SWELL Fishualization is aimed at enabling employees to gain insights into their working habits and encourage social interaction about healthy working, in order to improve well-being at work (T05). It provides a feedback screen in the form of a digital fish tank (see Figure 2.12), which is placed at a central location in the office. The primary sensor is currently a key-logging software that is installed on the user's computers. It captures key strokes, mouse movements and clicks together with information provided by the operating system: window titles, active applications, application switches, etc. Other sensors could also be coupled to add information on, e.g. heart rate, dominant facial expression or e-mail sentiments. Each fish in the Fishualization represents an individual employee. The speed of horizontal movement of a fish is determined by how fast the corresponding employee is interacting with their computer (number of clicks

---

<sup>5</sup>Chapter 8

<sup>6</sup>My contribution: set-up of the evaluation study.

and keystrokes) and the number of changes in direction per time unit represents the number of task or context switches per time unit. The y-position of each fish currently represents the (self-reported) energy level of the corresponding employee. 'Plants' at the bottom of the screen represent performed tasks, for example writing e-mail, editing document, browsing, or preparing presentation. The more people work on a task, the larger the plant.

Regarding our framework, the SWELL Fishualization is aimed at enhancing support from colleagues, thus addressing the working conditions. It tries to tackle stress in the middle of the stress chain, helping employees to cope with stress (secondary prevention). Its main basis is the Job Resources-Demands model (providing additional resources). It measures work characteristics and assesses the energy dimension of engagement by means of user input. With respect to behavioral change it helps with monitoring the current situation. Moreover, it fosters the motivation to take action by means of a playful approach and social influences.

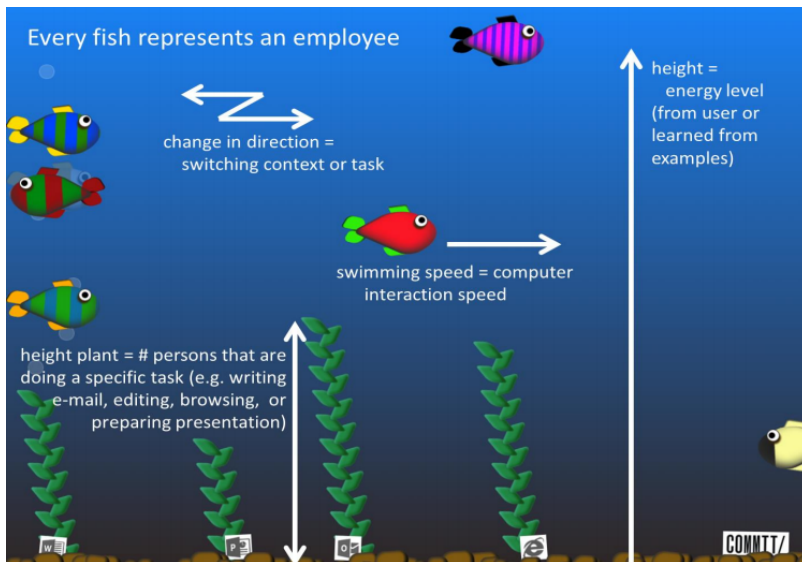


Figure 2.12: SWELL Fishualization. This screen is placed in the coffee corner and provides department feedback.

**Feedback on our ideas.** In the workshop described earlier, another subgroup of 6 participants (1 male, average age = 26.8) gave feedback on the SWELL

Fishualization feedback screen. The aspects that the group liked most were: 1) The design, the playful visualization. 2) That it combines several factors (team and individual, overview). 3) It makes curious and there are continuation options (conversation, signal, help). Moreover, they said that it is informal and light, makes the topic negotiable, stimulates to take action, and gives insights to help colleagues.

When asked what they would want to change or add, the main aspects mentioned were: 1) Trends over time (day, week, month). 2) Objectification, measuring relevant aspects. 3) The current system could build pressure (showing that you work hard). 4) Developing the system further as team (topics, focus, form). Moreover, they suggested adding more variables (heart rate, blood pressure, breaks, other applications), a coupling to actions, tips, suggestions, and help to bring about a dialogue (which questions to ask). Furthermore, they said that the prototype could provide positive appreciation or a prize, the possibility to design your own fish, and variations in the visualization (otherwise it becomes boring). Participants were concerned that the system may be confronting, and that it should not be too open (choosing on what you want to share with the team).

**Evaluation study.** We also evaluated the prototype in a real-world environment. The Fishualization trial at the Media and Network Services group at TNO ran for about 2.5 months (March - May 2014). The Fishualization screen (a large computer display) was placed in the coffee corner. A subset of 10 employees volunteered to couple their computer interactions and subjective input of their energy level to one of the fish. In order to measure the effects of the deployment of the Fishualization, all employees who use the coffee corner were asked to fill in pre- and post-questionnaires on personal awareness of working patterns and well-being at work, group awareness and interactions with colleagues. Furthermore, camera and microphone recordings were used to measure activity at the coffee corner, see Figure 2.13. To ensure privacy, only the number of detected faces, the amount of video motion and the average sound level were deduced and stored (no video or sound was stored). This data collection started 3 weeks before the Fishualization was turned on and continued during the trial, to compare activity in the coffee corner before and after deployment of the Fishualization.

30 employees filled in the pre-questionnaire and 14 employees filled in the post-questionnaire. (The subset of respondents did not differ significantly in their current level of well-being or how content they were about their well-being.) We used independent samples t-tests to compare the pre- and post-test results. A significant effect on the following item was found: “I am aware of

typical patterns in working behavior throughout the day or week (e.g. mailbox on Monday morning, project work after lunch...).” ( $p = 0.004$ ). Awareness of working patterns was higher in the post-test ( $M = 4.79$ ,  $SD = 1.626$ ) than in the pre-test ( $M = 3.27$ ,  $SD = 1.530$ ) (scale from 1: ‘not’ to 7: ‘very much’). Moreover, we found a significant effect on the item “I know how I can change my working behavior to gain a better level of well-being (e.g. becoming more productive, reducing stress...).” ( $p = 0.005$ ). Scores were higher in the post-test ( $M = 5.14$ ,  $SD = 1.231$ ) than in the pre-test ( $M = 3.9$ ,  $SD = 1.322$ ).

We can conclude that the Fishualization caused more personal awareness on working behavior and its relation with well-being among employees. However, we did not find significant effects on items related to group awareness and interactions with colleagues. In the further development of the Fishualization we should focus on fostering social interaction among colleagues more (e.g. by adding new functionality), as this may be a good buffer against stress. Moreover, most participants were enthusiastic about the Fishualization. A playful manner of feedback turned out to be engaging. Finally, we used sensor technology to quantify activity in the coffee corner, which shows the potential of new technology for experimental evaluation.

### 2.7.3 Providing tips - SWELL NiceWork e-coach

We here shortly present the SWELL NiceWork e-coaching app. (For details we refer to the original work presented in Wabeke (2014)<sup>7</sup>).

The SWELL NiceWork app is designed to provide coaching for short recovery breaks (T07). The app provides simple tips, 3 times a day, aimed at promoting well-being at work (see Figure 2.14). Various scientific articles, websites and magazines on well-being at work were reviewed to collect appropriate tips, which resulted in a list of 54 tips. Each tip does not take more than three minutes, and no special materials or specific locations are required. The recommended well-being tips are of different types: cognitive-behavioral, creative, physical exercises, food, journaling, relaxing, social, and time-management.

We found that different people had different preferences for tips (pilot study, in which 26 employees rated their preferences for the 54 tips). Therefore, a recommendation approach was chosen to adapt which tips are given to the specific user. A content-based predictor turned out to not be very accurate. This means that based on characteristics of the tip, such as the type, goal and focus, no clear preference profile could be made. A collaborative-based recommendation approach, however, proved to work well. This means that estimating the preferences for tips by aggregating ratings from peers with a similar rating behavior is

---

<sup>7</sup>My contribution: Supervisor of Master thesis internship.

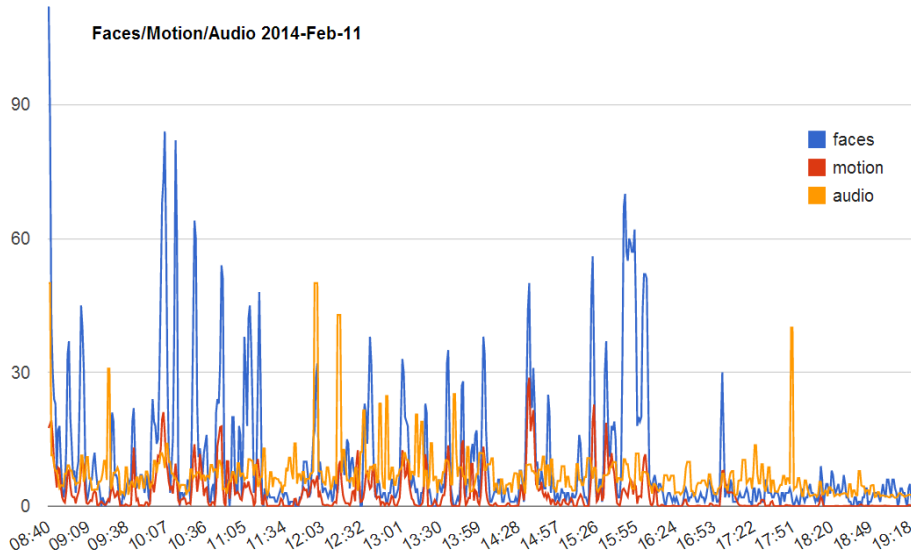


Figure 2.13: At the start of the working day (10-11) many faces are detected in the coffee corner. At lunch time (12), peaks in audio can be seen. In the afternoon, the coffee corner gets busy from time to time, until 4, when the coffee corner gets empty again. Such sensor recordings can be used to quantify the effect of the SWELL Fishualization: e.g. do people stay longer to watch at the screen, does the amount of audio increase (more people talking).

promising. An android app was implemented to provide well-being tips based on a recommendation engine. After each recommendation, the user can indicate whether he performed the tip and the system learns over time to give better tips.

Regarding our framework, the SWELL NiceWork e-coach is mainly aimed at supporting the work-rest balance. The app provides interventions with various goals: some are aimed at preventing the experience of stress (secondary prevention), and some on recovery from coping with high demands (tertiary prevention). The tips focus on personal factors or the working context. Its main basis is the Effort-Recovery model (focusing on recovery). It does not yet measure anything. The system does assesses whether the user has followed up a tip by means of user input. With respect to behavioral change it helps with taking action and learning new behavior by providing triggers and suggestions.



Figure 2.14: SWELL NiceWork app. Left: Intervention is provided. Right: Overview of provided interventions.

**Feedback on our ideas.** In the workshop described earlier, another subgroup of 6 participants (2 female, average age = 29.7) gave feedback on the SWELL NiceWork e-coaching app. The aspects that the group liked most were: 1) That the system adapts to the individual. 2) That the tips are really easy. Moreover, they said that it is a positive solution, which gets people out of their routine, and provides gradual behavioral change. The system has a great variation in tips in an accessible app, which is a relatively cheap solution to the work stress problem.

When asked what they would want to change or add, the main aspect mentioned was: 1) The user should stay in control of the system (e.g. help me today with...). Moreover, they suggested a need driven app (but what are the needs?), more personal advice, adding goals, tips about productivity, good timing of tips, a social community, and a coupling with sensor data. Participants expressed concerns regarding user acceptance, stopping to use the app and its reliability.

**Evaluation study.** To evaluate the NiceWork app with users, 35 employees tested the e-coach for 2 weeks. The first hypothesis was that knowledge workers have a positive attitude towards the e-coach. This hypothesis was confirmed in the user study. The number of followed-up tips was high (2 out of 3 per day)

and most participants agreed that it is pleasant to receive automatic notifications. The study also showed that three recommendations per day seemed a right amount of suggestions. Moreover, indicating whether a tip was followed-up and asking for a short motivation when a tip was rejected turned out to be a well-design method for providing feedback. Our second hypothesis was that tailored recommendations are followed-up more often compared to randomized suggestions. We did not find strong evidence for this hypothesis. Results show that our recommendation method, which provides tailored suggestions, did not substantially increase the number of tips that were performed compared to a method that provided randomized suggestions.

Furthermore, results show that only a few tips were not followed-up, because the tip was disliked (13%). Instead, tips were mostly rejected, because the moment of recommendation was somehow inappropriate (wrong timing: 46%, tip not relevant: 15%, not at work: 14%). This finding suggests that future e-coaches may increase their effectiveness by recommending tips at appropriate times. Using sensor information to ensure that tips are suggested just-in-time, was the most important personalization method that needed to be further explored. (This was done in Schendel (2015).) Moreover, we demonstrated that technology can be used to investigate the effects of an intervention, i.e. via the app we directly investigated how many interventions were said to be followed up, and we directly asked for reasons for not following up a suggestion.

#### **2.7.4 Providing support - SWELL HappyWorker system**

We here shortly present the SWELL HappyWorker system. (For details we refer to the original work presented in Sappelli and Verberne (2014) and Sappelli (2016).)

One method for reducing the demands on a knowledge worker is by reducing information overload. The HappyWorker prototype recommends documents and webpages that are relevant to the current project context. The sensor used is computer logging. To enable context recognition, a new networked model was implemented, the CIA model. The prototype provides a screen where the user can see which active project the software has detected in real-time, and he can access recommended documents and webpages. In the list of recommendations, the documents that are most highly activated in the CIA model end up. As alternative, two other approaches to context-aware document recommendation were also implemented (a content-based recommender with contextual pre-filtering (CBR) and a just-in-time information retrieval system (JITIR)).

Regarding our framework, the SWELL HappyWorker system is aimed at providing work support, thus addressing the work itself. It tries to tackle stress

in the beginning of the stress chain (diminishing demands) with the aim of addressing the stressor itself (primary prevention). Its main basis is the Job Resources-Demands model (diminishing demands). It measures work characteristics. With respect to behavioral change it helps with taking action and learning new behavior by means of a supportive environment.

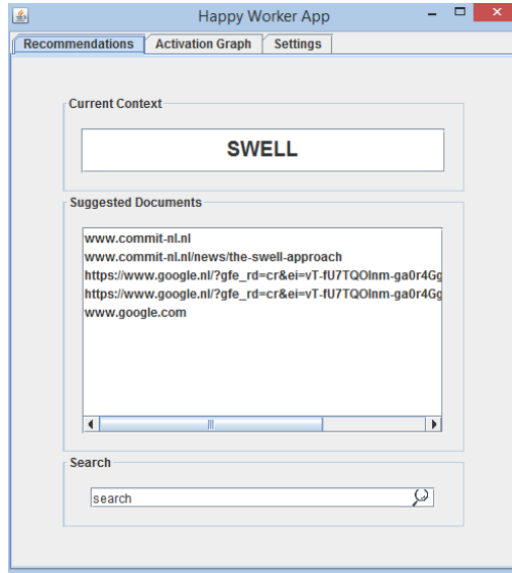


Figure 2.15: SWELL HappyWorker app, showing: the active project that the software detected and recommended documents.

**Evaluation.** A technical evaluation of the system was performed. The SWELL dataset was used, in which 25 students wrote reports and prepared presentations. The collected sensor-data from the installed key logger and the file history were used for topic inference. A ground truth topic label was available for each block of computer activity to evaluate the system (labeled by human annotators; e.g. 'road trip in the USA', 'Napoleon'; 8 labels in total; see Sappelli, Verberne, *et al.* (2014)).

It was found that the new CIA model was effective in classifying the user's context with an average accuracy of 64.85%. The different context-aware document recommendation methods proved effective in different evaluation tasks. The new CIA model was best at suggesting documents that the user indeed opened later and provided variety in its suggestions. JITIR was best at suggest-



ing documents whose text was later included in documents that the knowledge worker wrote. CBR was best at only suggesting documents relevant to the topic at hand, therefore preventing distractions.

To conclude, it is technically feasible to recognize the knowledge workers' current working context and to suggest relevant documents to the topic at hand, while requiring little user effort. In future work, user studies should be performed, evaluating the system with knowledge workers at work. The focus should be on the user experience and effects on demands posed upon the knowledge worker. Moreover, the scalability of the algorithms to a real world situation should be investigated.

### 2.7.5 Conclusions: Evaluation of prototypes

In this section we presented the general SWELL functionality and described four prototypes and their evaluation. (Within the SWELL project also other prototypes were developed and evaluated, e.g. ([Brighter 2015](#)).) To sum up:

- The SWELL Workload Mirror is aimed at providing insights in stress sources. It uses various sensors to give employees feedback. From the evaluation study we gained many valuable insights on potential hurdles to use such a pervasive system. The most important point of attention is the user's privacy. We are continuing our work on this prototype by taking the insights on user concerns into account, e.g. by applying privacy by design.
- The SWELL Fishualization is aimed at fostering interactions among colleagues. It provides a feedback screen in form of a fish tank. The evaluation study showed us that a playful manner of feedback causes a lot of interest and curiosity among employees. We are continuing our work on this prototype by testing the Fishualization in other companies, adding more types of sensor information and improving the look and feel of the fish tank.
- The SWELL NiceWork e-coach is aimed at providing interventions. It provides simple quick tips that can be performed during work to improve well-being. What the evaluation study showed us is that employees valued the app and on average 2 of 3 tips per day were followed up. The area in which most improvement can be gained is the timing of tips. We are continuing our work on this prototype by doing additional research on appropriate timing of well-being tips.

- The SWELL HappyWorker system is aimed at providing support. It helps the user to find information that is relevant to his current working context. The evaluation study showed us, that the current working topic can be inferred from computer (interaction) data. Moreover, the system can produce relevant content suggestions in real-time. The next step would be to evaluate and develop the HappyWorker system further with potential end users.

In general we can say that we made working implementations of some pervasive technologies for improving well-being at work. Our evaluation until now was mainly aimed at user experience and testing underlying technologies. The evaluation yielded several additional requirements for our system. Moreover, we showed how technology can be used to investigate the effects of an intervention. In further research we should also evaluate whether the prototypes have the expected positive effect on employee's well-being at work. From our small scale pilot studies we got some first insights, but ideally the systems are evaluated with in a much larger field test.

## **2.8 Conclusions and Discussion**

In this work, we combined stress and intervention theory with knowledge of technological possibilities and input by users, to design a pervasive system that helps knowledge workers to improve well-being at work.

### **2.8.1 Conclusions**

We answered the following research questions:

1) Which concepts are relevant with respect to well-being at work? We found that the relationship that people have with their jobs can be described as a continuum between engagement and burn-out. Engagement is characterized by energy, involvement and efficacy or absorption. Biology describes more short term effects of stress. A stressor causes a particular perception of the stressor in the individual. This can lead to acute physiological stress responses and, on the long run (due to lack of recovery) to long term physical, cognitive, emotional and behavioral stress consequences.

2) Which person, work and context conditions can lead to negative stress? There are no specific personal, work or context conditions that generally lead to stress. Work becomes stressful when high demands are combined with: insufficient resources; little rewards; little recovery; or an environment that mismatches with personal characteristics. The most useful models for developing

technology based interventions are the Job Resources-Demands model and the Effort-Recovery model. We presented several ideas to diminish demands, enhance resources or help with recovery.

3) How can sensors be applied to automatically infer stress and the context in which it appears? We can use technology to sense work characteristics (e.g. tasks and topics worked on), measure acute physiological stress responses in the body (e.g. HRV), or assess cognitive, emotional and behavioral effects of stress (e.g. sleep duration). The user study showed that users are only interested to collect data that is necessary for supporting their specific goal, so the system should be configurable.

4) Which interventions can be provided by means of pervasive technology to help a knowledge worker improve his well-being at work? Either the stressor can be addressed (e.g. work support), the short-term stress reactions (i.g. enhance coping), or long-term stress consequences (e.g. improve recovery). Moreover, different areas can be addressed: the work itself, the working conditions, personal factors or private circumstances. For behavioral change, the system should support the employee to: monitor the current situation and identify problems; set change goals and plan actions; and help the employee to take action and learn new behavior. Suitable behavioral change techniques should be used to address the motivation, ability or trigger to take action (e.g. feedback, self-monitoring, risk communication and reminders to action).

We presented the resulting general framework in which we related several relevant theories. This framework can be used by other researchers to design pervasive systems that address well-being at work.

Finally, we described the envisioned SWELL system, and core functionality that was identified. We also presented some built prototypes. The SWELL Workload Mirror provides an activity and workload overview, designed to find stress sources. The SWELL Fishualization provides department wide feedback for peer support, designed to improve coping. The SWELL NiceWork e-coach provides well-being tips, designed to improve coping or recovery. Last, the SWELL Happy Worker system provides personalized search, designed to support work. All in all, we demonstrated the (technological) feasibility of our ideas. First evaluations with users were positive and provided further insights to refine the systems.

### **2.8.2 Discussion**

The biggest challenge in developing our comprehensive and practical framework was the vast amount of available concepts and models regarding well-being at work. We consulted experts in the field. We finally, had to make choices on what

concepts and theories to include. Our selection may reflect our specific scoping. We focused on providing a general and simple overview, combining different areas of research.

Another big challenge in this respect was relating concepts of different fields to each other. These concepts differ in their level of abstraction: Organizational Psychology provides the most high-level terms, i.e. the relation between resources vs. demands, or recovery. Biological theories provide more low-level terms, i.e. physiological stress responses in the body. Our aim was to make several of these aspects quantifiable by means of sensors. This means, translating these concepts into even more low-level terms, i.e. a specific variable to infer, the information necessary and specific sensors.

Besides the high-level vs. low-level continuum, there is also a temporal continuum, from short-term stress to developing a burn-out. In traditional approaches with questionnaires, mainly long-term aspects are assessed. Sensing, however, enables real-time measurements in real world work settings. We aimed to translate relevant aspects identified based on theories, into variables that are measurable at the workplace.

The resulting general and pragmatic framework provides a structure to develop pervasive technology for improving well-being at work. We noticed that far more diverse technology based interventions can be developed, than initially assumed. The theoretical foundation gave many different pointers of how well-being at work can be improved: from coaching during work, over fostering social support, to addressing recovery after work. Besides the ideas and prototypes presented here, many more (technological) solutions can be developed based upon this general framework. To mention just a few examples: teaching coping in an online course, building a social network for peer support, or enhancing recovery by letting people play a computer game.

Moreover, we made working implementations of some pervasive technologies for improving well-being at work. Our evaluation until now was mainly aimed at user experience and testing underlying technologies. Further research should evaluate whether the prototypes have the expected positive effect on employee's well-being at work. From our small scale pilot studies we got some first insights, but ideally the systems are evaluated with in a much larger field test.

As a final note, we need to be cautious to put responsibility for managing work stress at the individual level. Certainly the company and management also play a role. Therefore, an intervention provided one on one at an individual by means of a pervasive system, is ideally part of a larger intervention program. In case many employees struggle with similar problems, a department wide intervention may be more effective. Furthermore, specific problems at work may not be solvable by the employee himself. In this case, the management or

organization may need to be approached.

### 2.8.3 Identified research challenges and opportunities

All in all, we think a pervasive system aimed at an individual's abilities to cope with stress and improve well-being at work poses many new opportunities. A system used real-time during work can provide much valuable information on work stress. Moreover, employees can be empowered to self-manage their well-being at work by means of tailored interventions. Throughout our work we encountered several challenges and opportunities for further research in several categories:

1. **Multi-disciplinary, theory and data-driven research and development.** New technology brings new possibilities. The now very abstract models can be more refined to include directly measurable concepts and new types of support. New technology can also be used to directly evaluate the success of an intervention. Sensors can be used to investigate in how far interventions are indeed followed up (e.g. whether users take a break or become physically active after a suggestion by an e-coach). Moreover, the effects of an intervention can be measured (e.g. whether provided information support indeed decreased mental effort and stress). Technical experts and social scientists should aim to work together. It is therefore necessary that the experts understand each others domains well, which is challenging.
2. **Interpreting personal sensor data.** Sensor data is relatively easy to collect, the challenge is making sense of this data. We should investigate which behavior is indicative of stress during work and how these can best be captured by means of unobtrusive sensors. People differ in their (work) behavior, so there is a need to build personalized models This brings methodological challenges, e.g. how to instantiate a model for a new user.
3. **Relation between measurable aspects and long-term stress consequences.** In future work, the relation between subjective experience based upon our own feelings and objective measures based on objective data should be investigated. Can objective measurements help us with detecting stress? Ideally, a system would be able to give a warning in case it predicts that the current behavioral pattern will cause long-term problems. Therefore, research should be done on how longitudinal patterns in sensor data relate to long-term stress consequences and burn-out.

4. **Combining strengths of human and computer.** Ideally, the strengths of a computer (e.g. being objective or persistent) and the strengths of a human (e.g. being good in interpretation) should be combined. The role of the system and the user should be clear. The most suitable manner for pervasive technology to interact with an employee is a challenging question for human-computer interaction research. Issues of control are important. The system needs to interact in a way that provides support, while not irritating the user.
5. **Privacy.** The success of pervasive systems collecting context data depends on the acceptance by users. A system that collects personal data raises many privacy questions. Therefore, privacy should be integral part of the design process (e.g. doing a Privacy Impact Assessment or implementing Privacy by Design).
6. **Ethics.** Measuring and trying to change the behavior of individuals poses all kinds of ethical questions. Is it acceptable to monitor and change the behavior of an employee? It is difficult to predict how such new pervasive e-coaching systems will be perceived and used (or even misused) when applied in real-world work settings.



## The SWELL knowledge work dataset for stress and user modeling research

*This chapter describes the new multimodal SWELL knowledge work (SWELL-KW) dataset for research on stress and user modeling. The dataset was collected in an experiment, in which 25 people performed typical knowledge work (writing reports, making presentations, reading e-mail, searching for information). We manipulated their working conditions with the stressors: email interruptions and time pressure. A varied set of data was recorded: computer logging, facial expression from camera recordings, body postures from a Kinect 3D sensor and heart rate (variability) and skin conductance from body sensors. The dataset is available for access by the scientific community and not only contains raw data, but also preprocessed data and extracted features. Moreover, the participants' subjective experience on task load, mental effort, emotion and perceived stress was assessed with validated questionnaires as a ground truth. The resulting dataset on working behavior and affect is valuable for several research fields, such as work psychology, user modeling and context aware systems. We use this dataset for our analysis of work behavior and for inferring work stress presented in Chapters 5 and 6.*

---

This chapter is based on Koldijk, Sappelli, Verberne, *et al.* (2014). "The SWELL Knowledge Work Dataset for Stress and User Modeling Research". Published in: Proceedings of the International Conference on Multimodal Interaction (ICMI 2014).



### 3.1 Introduction

Nowadays most work involves computer usage and information processing. People that use and produce information as their main task are called knowledge workers. They typically experience all sorts of demands during their work days, such as several tasks that need to be finished before a deadline (high workload, temporal demand). For this they need to combine different information sources, for example from the internet (requiring mental effort). Incoming emails may be an important source of distraction during a task (potentially causing frustration). In case people feel they cannot handle the demands posed upon them, they can experience stress (Demerouti, A. B. Bakker, Nachreiner, *et al.*, 2001). Stress is a broad concept referring to psychological and biological processes during emotional and cognitive demanding situations. We follow a pragmatic approach and define stress in terms of: (1) the task load, which poses demands on the worker, (2) the mental effort, which the worker needs to handle a task and (3) the emotional response to a task. In this chapter, we focus on short term effects of stressors that can be measured within a 3 hour work session.

Stress is a well-known experience in our connected environments. Ruff (2002) speaks of ‘hurry sickness’ as “the belief that one must constantly rush to keep pace with time” and ‘plugged in compulsion’ as “the strong need to check mail and the internet to stay in touch”. Mark, Gudith, and Klocke (2008) investigated the cost of interruptions and came to the conclusion that “after only 20 minutes of interrupted performance people reported significantly higher stress, frustration, workload, effort and pressure”. Stress from time to time with enough room for recovery is no problem (Demerouti, A. B. Bakker, Geurts, *et al.*, 2009). However, when stress builds up this can be a danger to well-being, in the worst case resulting in burn-out.

In our project SWELL (Smart Reasoning for Well-being at Home and at Work)<sup>1</sup> we aim to develop ICT tools that help knowledge workers to cope with stress and gain more well-being at work (Koldijk, 2012). We want to interpret recordings in the office real-time in terms of stress and the context in which it appears. Based upon this information, we aim to develop coaching software that can help knowledge workers to gain a more healthy work style. Moreover, we want to develop smart information support tools that assist the knowledge worker in handling the large amount of (incoming) information he has to work with. In this way, we extend traditional approaches (e.g. questionnaires or department wide interventions, Koppes *et al.* (2012) and Kraan *et al.* (2000)) by empowering individual users to self-manage their own well-being.

---

<sup>1</sup><http://www.swell-project.net>

To be able to develop the ICT tools we envision, research communities like work psychology, user modeling and context aware systems are in need of a good dataset. This dataset should ideally have the following characteristics: Data should be recorded in a realistic office setting. Stressors should be manipulated in a systematic way and subjective experience should be assessed with validated questionnaires, to be able to investigate the effects of stressors. A multimodal set of sensors from different research fields should be used, to enable multidisciplinary research. The focus should lie on sensors that are readily available in office settings, to make the to be developed system usable outside the lab. To our knowledge no such dataset existed.

In this chapter we present a newly collected rich dataset which has these characteristics. Our dataset overcomes three drawbacks that are typically observed in related work:

- Instead of a rather artificial task, participants perform natural office work with systematically manipulated stressors.
- Instead of expensive and/or obtrusive equipment, we decided to combine a variety of sensors that can easily be deployed in real-world office settings.
- Instead of only collecting data for our own use, the dataset has been curated and is archived sustainably by the DANS archive of KNAW at <http://persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-kwrv-3e>. In this way the scientific community can use it for benchmarking of techniques and algorithms. Not only raw data is provided, but also data in preprocessed and interpreted form.

With our new dataset, we aim to bring research on psychology and computer science together. With this dataset, research questions from several fields can be answered, for example:

- Work psychology: What effect do stressors like time pressure have on the working behavior of knowledge workers? What is the effect of an incoming email? What effect do stressors have on subjective experience of task load, mental effort, emotion or perceived stress? What is the relation between what people mean when they feel 'stressed' and the concepts of arousal, mental effort and valence? Do we see effects of stressors in physiological sensor data?
- User modeling: Can we estimate the mental state or emotion of knowledge workers from unobtrusive sensor data? Do knowledge workers show

particular affective expressions during computer work? Are there typical facial expressions or postures that are indicative of mental effort, high workload or stress?

- Context aware systems: Can we automatically determine the task or topic someone is working on? Is there a relation between stress and the context in which it occurs? Can we filter irrelevant emails? Can we make information retrieval more context aware?

In this chapter we mainly focus on work stress and user modeling. For more details about the data regarding context recognition and information support, we refer to Sappelli, Verberne, *et al.* (2014).

This chapter is structured as follows. We first present some related work (Section 3.2). We then outline our experimental setup in which the dataset was collected (Section 3.3). We describe the dataset in detail in Section 3.4. Some example analyses are presented in Section 3.5. We finish with a Discussion and Conclusion (Section 3.6 and 3.7).

## 3.2 Related Work

In this section we present some related research on work psychology and user modeling, in which sensor data is used to estimate stress, mental or affective states. We specifically address the type of sensors used, the context in which data has been collected and the kind of inferences that have been made. This gives a theoretical framework for research on our collected dataset.

In work psychology, questionnaires are commonly used to get insight in the general working experiences (e.g. Zapf (1993)). Advances in sensing, as well as the quantified self movement make it possible to extend such an approach with on-site measurements.

Work in the area of affective computing investigates the possibility of inferring stress and emotion from sensor data. Most often, physiological sensors are used and data are collected in experimental environments. In research by Riera *et al.* (2012), for example, electroencephalography (EEG) and facial electromyography (EMG) data were collected. The authors show that EEG and EMG can be used for monitoring emotion (valence and arousal) and stress. Despite its great potential, we think deploying EEG in a daily office setting is not yet realistic. Other common measurements in stress research are pupil diameter and electrocardiogram (ECG). Mokhayeri, Akbarzadeh-T, and Toosizadeh (2011), for example, collected such data in context of the Stroop color-word (SCW) test. They state that pupil diameter and ECG have great potential for stress detection. However, the question that arises is: can we also make an estimate of

affective and mental states outside the lab? We see some potential for ECG measurements, with the rise of wearable sensors, which are becoming more and more integrated into devices as watches and bracelets. But, besides measuring the physiological stress response directly, we also see great potential in measuring outward characteristics, such as facial expressions, postures or computer interactions as indicators for the user's state.

In related work, facial expressions are widely used for inferring emotions. The data are often recorded while emotions are induced in participants. The publicly available multimodal dataset described by Soleymani *et al.* (2012), for example, was collected in context of watching emotion inducing video clips and consists of: face videos, audio signals, eye gaze data and physiological signals (EEG, ECG, GSR, respiration amplitude, skin temperature). Although this dataset is very interesting, emotions in a daily computer work context are probably less intense than the valence or arousal experienced during watching a movie clip. An interesting question is whether people show facial emotions during computer work, and whether their facial expressions are indicative of mental states. Preliminary results by Dinges *et al.* (2005) suggest that high and low stressor situations could be discriminated based on facial activity in mouth and eyebrow regions.

Regarding postures, Kapoor and Picard (2005) present research in which posture data was collected together with facial expressions and computer information while children solved an educational computer puzzle. Sensors in the chair were used to extract posture features (like leaning back, sitting upright) and activity level (low, medium, high). Posture information yielded the highest unimodal accuracy (80%) for estimating interest (vs. uninterest). Performance was further improved by adding facial expression and computer information. We conclude that posture information and movement are an interesting source for estimating the users' mental state. We see potential for posture measurements in the office, as with the Kinect recently an affordable 3D camera with skeleton detection has entered the market.

Finally, in some research, stress or emotions are estimated from computer interaction data. Vizer, Zhou, and Sears (2009), for example, investigated the effect of stress on typing patterns. Participants first performed a mentally or physically stressful task (e.g. remembering digits or exercising) and were then asked to write an email. Results indicate that stress can produce changes in typing patterns. This makes computer logging a valuable sensor for user state modeling. We think not only typing patterns, but also more general computer behavior might be indicative of mental states, like the amount of window switching, number of typos or time spent browsing.

Besides inferring stress or particular mental or affective states, the context in

which they appear can be interesting. Computer interactions give rich insight in the user's current working behavior. Research by Koldijk, van Staalduinen, Neerincx, *et al.* (2012) shows that it is possible to infer the task someone is working on from computer interaction data. Moreover, one could add analysis of contents worked on.

To conclude, research from various related fields shows the potential of using sensors for estimating stress, mental and affective states and the context in which they appear. In each field, a particular setup of sensors is used. We decided to combine several of these in our unique dataset: computer interactions, video for facial expressions, Kinect 3D for postures and body sensors for heart rate and skin conductance.

### 3.3 Data Collection Context

In this section we present the experimental setup that was used to collect data.

#### 3.3.1 Design

In our experiment we manipulated the conditions under which our participants worked:

- Neutral: the participant was allowed to work on the tasks as long as he/she needed. After a maximum of 45 minutes the participant was asked to stop and told that enough data of 'normal working' was collected.
- Stressor 'Time pressure': the time to finish all tasks was 2/3 of the time the participant needed in the neutral condition (and maximally 30 minutes).
- Stressor 'Interruptions': 8 emails were sent to the participant during the task. Some were relevant to one of the tasks, others were irrelevant. Some emails required a reply, others did not. Examples are: "Could you look up when Einstein was born?" or "I found this website with lots of nice pictures for presentations."

All participants worked under all 3 conditions. The neutral condition was always the first condition, in order to collect an uninfluenced baseline of normal working. The order of the two stressor conditions was counterbalanced, see Figure 3.1. The within-subject design included relaxation breaks to start each condition in a well-rested state.

Order		Block1	Q		Block2	Q		Block3	Q
A	R	Neutral		R	Stressor interruptions		R	Stressor time pressure	
B	e			e			e		
	l			l			l		
	a	Neutral		a	Stressor time pressure		a	Stressor interruptions	
	x			x			x		

Figure 3.1: Design. For 13 participants order A was used, for 12 participants order B.

### 3.3.2 Tasks

The participants performed knowledge worker tasks on a desktop computer in a controlled lab setting. We asked them to write reports and make presentations on predefined topics (in English). We selected 6 topics on which people with various backgrounds could work:

- 3 opinion topics: Experience and opinion about ‘stress at work’, ‘healthy living’ and ‘privacy on the internet’.
- 3 information topics: ‘describe 5 Tourist attractions in Perth (West Australia)’, ‘plan a coast to coast road-trip in the USA’ and ‘write about the life of Napoleon’.

Some detail on what to include in the report was also given. Participants were allowed to look for information on the internet and use documents that we previously stored on the computer. This setting is typical for knowledge work as available information can be combined with the worker’s own input in a coherent way, with the purpose of generating a new information product. During the task the email program Outlook was running and participants were told to make use of information from incoming emails and reply when necessary. In this way, a realistic office work scenario was created.

We wanted to ensure that the participants worked on the tasks seriously. Therefore we told them that it was important to finish all required tasks for receiving the full subject fee. Moreover we told them that they would have to give one of the prepared presentations. After the experiment, we debriefed all participants and informed them that they did not need to give a presentation and would get the full subject fee.

### 3.3.3 Procedure

To be able to record stress responses as a result of our experimental manipulations, we instructed the participants to not smoke or drink caffeine 3 hours prior to the experiment, as these are possible confounders. Before the experiment started, the experiment and recordings were explained and all participants signed a consent form to confirm that the recorded data may be used for research purposes. Body sensors were applied and while the experimenter checked the recordings, the participant read the experiment instructions and filled in a general questionnaire.

The experiment was divided into three blocks for the different stressor conditions, each taking approximately one hour. Each of the experimental blocks started with a relaxation phase of about 8 minutes (which is typical for stress research) in which a nature film clip was shown. Then the participants received instructions on the tasks to work on. In each block the participants were provided with 2 of the 6 topics, which were randomly selected from the list, in such a way that always an opinion topic was combined with an information topic. The participants were instructed to write 2 reports, one on each topic, and make 1 presentation on one of the topics (participants could choose the topic). To prevent learning effects, the participants were provided with different topics in every block. In both stressor conditions, participants were provided a count-down clock for showing them the remaining time.

After completion of the tasks, the participants were asked to fill in a questionnaire about the current block. This procedure of relaxation, tasks execution and questionnaire was then repeated for block 2 and 3 (see Figure 3.1). Between the conditions the subjects were allowed a short break and the total experiment took about 3 hours. After the experiment the participants were debriefed.

### 3.3.4 Apparatus

Participants performed their tasks on a computer (Dell Latitude E6400) with Windows 7 Professional with a 17 inch screen and mouse and keyboard (see Figure 3.2). Office 2010 was installed, which the participants used for email (Outlook), report writing (Word) and making presentations (Powerpoint). As a browser, Internet Explorer was used with Google as default search engine. The start page of Internet Explorer was [www.google.nl](http://www.google.nl).

### 3.3.5 Subjective Ratings

To collect a ground truth of the subjective experience after each block, we used a combination of validated questionnaires. Task load (in terms of mental de-



Figure 3.2: Experimental set-up.

mand, physical demand, temporal demand, effort, performance and frustration) was determined with the ‘NASA-Task Load Index’ (Hart and Staveland, 1988). Mental effort was assessed with the ‘Rating Scale Mental Effort’ (Zijlstra and van Doorn, 1985). Emotion response (in terms of valence, arousal and dominance) was determined with the ‘Self-Assessment-Manikin Scale’ (Bradley and Lang, 1994). Moreover, we asked participants to report their perceived stress on a visual analog scale from ‘not stressed’ to ‘very stressed’ (10 point scale).

Furthermore, we asked the participant’s to fill in the ‘Internal Control Index’ questionnaire (Duttweiler, 1984). People with an internal locus of control tend to praise or blame themselves, whereas people with an external locus tend to praise or blame external factors. This might be of influence on participants’ stress perception or behavior. Moreover, participants were asked to rate their interest in the topics, as well as how difficult they found it to write a report or make a presentation on a topic on a 7-point Likert scale (from ‘not interesting / difficult’ to ‘very interesting / difficult’).

### 3.3.6 Sensors

**Computer logging.** Computer interactions were logged with the key-logging application uLog (version 3.2.5, by Noldus Information Technology), which ran as a background application on the users’ computer.



**Video.** Video recordings of the participants' face and upper body were made with a high-resolution USB camera (iDS uEye UI-1490RE, 1152x768) which was positioned below the participants' monitor. The AVI files from the USB camera were further analyzed using the facial expression analysis software FaceReader (version 5.0.7 RC 4.5 (Beta)). An additional webcam (Philips SPC 900NC, SVGA resolution) was placed above the participants' monitor.

**Kinect 3D.** The participants body posture was recorded with a Kinect (for Windows, model 1517) depth camera. The camera was placed in front of the participants at a distance of about 2 meters, such that their whole body, including their legs under the desk were visible (see Figure 3.2). Besides 3D depth video, Kinect also recorded normal RGB video. Recordings were made with Kinect Studio (v1.7.0), which resulted in xed-files. From the recorded Kinect data the depth image and information on the skeletal model were extracted using the Windows Kinect SDK (v1.7). We smoothened the data with several predefined filters.

**Body sensors.** ECG was recorded using a Mobi device (TMSI) with self-adhesive electrodes. The electrodes were placed across the heart, one below the participants right collar-bone, the other left below the chest, with a grounding electrode below the left collar-bone. Some preprocessing was programed into the recording software Portilab2. To record skin conductance, Mobi was used with finger electrodes. These were fixed with Velcro tape around the lower part of the thumb and ring finger of the participant's non-dominant hand. Recording frequency was 2048 Hz. All signals (ECG and skin conductance, raw and preprocessed) were stored together in S00-files.

**Additional Lab Recordings.** The lab's ceiling camera and microphones were used for making recordings of the lab during the whole experiment, as well as a screen capture of the participant's screen. The video files are encoded in AVI-format with a codec specific to the labs recording software (GeoVision's CCS5). Audio is encoded in separate wav-files.

### 3.3.7 Participants

25 students participated in our experiment, of which 8 were female and 17 male. The average age was 25 (standard deviation 3.25). Most participants were native Dutch. They were interns from TNO and students from Delft University of Technology who were approached by advertising. Since these interns and students are experienced in handling (large amounts of) information for their courses, and often use computers as their most important tool, they are assumed to be representative of knowledge workers. Additionally, they are experienced with the knowledge worker tasks we have chosen: writing reports and preparing

Table 3.1: Our dataset contains data from 25 participants (3 hours each). The listed raw and preprocessed sensor data, as well as a feature dataset (aggregated per minute) will be made available.

Type	Available raw & preprocessed data	Available features (#)
Computer interactions	uLog output <sup>2</sup> (i.e. xml-logs of all computer events) Parsed data (i.e. txt-file with timestamped data)	Mouse (3) Keyboard (7) Applications (2) (for details see Table 3.2)
Facial expressions	FaceReader output <sup>3</sup> (i.e. txt-logs with facial information) Parsed data (i.e. txt-file with timestamped data)	Head orientation (3) Facial movements (10) Action Units (19) Emotion (8)
Body postures	Joint coordinates from Kinect SDK (i.e. txt-file with timestamped data) Angles of the upper body (i.e. txt-file with timestamped data)	Distance (1) Joint angles (10) Bone orientations (3x11) (as well as stdv of the above for amount of movement (44))
Physiology	Data from Mobi <sup>4</sup> (i.e. S00-files with raw & filtered signals)	Heart rate (variability) (2) Skin conductance (1)

presentations. The participants received a standard subject fee for their participation in the experiment.

To assess whether the participants worked on the tasks seriously, we checked the quality of the written reports and presentations. As the quality was satisfactory, none of the subjects needed to be excluded from the corpus. Of the participants, 2 were left handed and 8 wore glasses (which could be of concern for the software analyzing facial expressions). Results of our pre-questionnaire showed that none of the participants indicated to have a heart disease or take medicine which could have influenced their heart rate. About half the participants indicated that they were physically active before the experiment as they came by bike. 4 participants indicated that they had experienced stress prior to the experiment. None of the participants smoked, drank caffeine or alcohol 3 hours prior to the experiment. The participants scored on average 3.67 on the internal control index (scale from 1 to 5, with higher scores indicating more internal control; stdv = 0.29).

### 3.4 Dataset

In this section we present the public SWELL-KW dataset in more detail. We collected data from the following sensors: computer logging, video, Kinect 3D and body sensors. Handling this data requires expertise in different fields. We preprocessed this data to get an aggregation of computer interactions, extraction of facial expressions, postures, heart rates and skin conductance levels. We finally aggregated this data into features per minute. For an overview of all available data see Table 3.1. We now first describe the available fully preprocessed and aggregated feature data. Then we describe the available raw data and preprocessing.

#### 3.4.1 The Feature Data

The feature dataset contains our completely preprocessed data, aggregated per minute, for all 25 participants. It contains the following features: 12 computer interaction features, 40 facial expression features, 88 body posture features and 3 physiology features as listed in the right column of Table 3.1. The feature dataset is annotated with the conditions under which the data was collected. Per participant three times 6 minutes relaxation data are included, ca. 45 minutes of working under normal conditions, ca. 45 minutes working with email interruptions and ca. 30 minutes working under time pressure.

Moreover, we provide the scores on our questionnaire items as ground truth for the subjective experience in each condition, see Table 6.2. As 25 participants each rated 3 conditions, this yields 75 ratings in total.

#### 3.4.2 The Raw Data and Preprocessing

Besides the completely preprocessed and aggregated data, we also provide some raw data and files resulting from our preprocessing, as listed in the middle column of Table 3.1.

**Computer logging.** The computer logging software recorded detailed timestamped information in XML format about each computer event. Examples of computer events are mouse clicks, mouse scrolls and application changes. Moreover we parsed the files and printed them in a more intelligible timestamped table format, which will also be made available. Finally, we computed several relevant mouse, keyboard and application characteristics per minute (listed in Table 3.2), which are contained in the feature dataset.

---

<sup>2</sup><http://www.noldus.com/human-behavior-research/products/ulog>

<sup>3</sup><http://www.noldus.com/human-behavior-research/products/facereader>

<sup>4</sup><http://www.tmsi.com/en/products/mobi.html>

Table 3.2: Computer interaction features (aggregated per minute).

Type	Feature	Description
Mouse	MouseActivity	Number of all MouseEvents
	LeftClicks	Number of left clicks
	MouseWheel	Number of mouse wheel scrolling
Key-board	KeyStrokes	Number of all KeyEvents
	ShortcutKeys	Number of shortcut keys (Ctrl+ c/x/v/z/s/a; Shift+Tab)
	DirectionKeys	Number of direction keys (arrow left/right/up/down)
	Characters	Number of characters (a-z)
	CharactersRatio	#characters divided by #keyStrokes
	ErrorKeys	Number of error keys (Backspace, Delete, Ctrl+Z)
	ErrorKeyRatio	#errorKeys divided by (#characters + #spaces)
Appli- cations	AppChanges	Number of application changes
	TabfocusChange	Number of tab focus changes

**Facial expressions from video.** We do not include the fully recorded videos in our dataset to keep our participants anonymous. Instead, we provide data files with the analysis of facial activation. These were extracted from the video per timeframe using the software FaceReader. The characteristics that are included in the dataset are: quality, estimates on the orientation of the head, some global features of the face such as looking direction and the amount of activation in several facial action units. Moreover, FaceReader provides an estimate of the subjects emotion, which is also available in our dataset. We parsed these files to get a more intelligible timestamped table format, which will also be made available. Besides data per video frame, we also calculated averages per minute for all characteristics (see Table 3.1), which are contained in the feature dataset.

**Body postures from Kinect 3D sensor.** We do not include the recorded 3D Kinect files in our dataset to keep our participants anonymous. Instead, we provide data files with analysis of the participant’s body posture per timeframe. These were extracted from the 3D Kinect recordings using the Kinect SDK. By fitting the Kinect skeletal model (see Figure 3.3, left), we got coordinates of all body joints per frame. This data will be made available. We further used these joint coordinates to determine joint angles between bones of the upper body, for example the angle between the upper and lower arm. Moreover, we determined bone orientations of the upper body relative to the x, y and z axis (see Figure 3.3, right), for example the angle between the left shoulder and the up

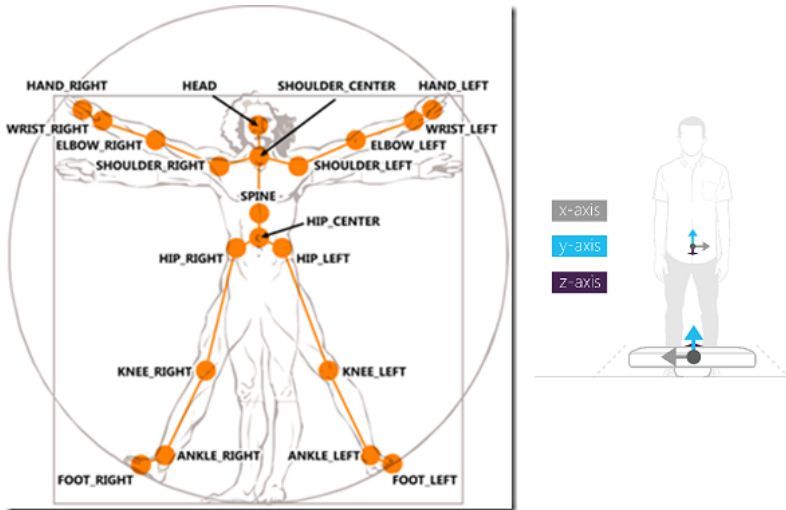


Figure 3.3: Kinect joints and Kinect orientation from Kinect SDK. Images from: <http://msdn.microsoft.com>

pointing y axis.<sup>5</sup> This information on angles per frame will be made available. From the depth image the average distance of the user was also determined. Finally, we determined average angles per minute, which are contained in the feature dataset. We also calculated standard deviations for each minute, to determine features that indicate the amount of movement and changes in joint angles. These are also contained in the feature dataset.

**Physiology from body sensors.** We provide raw and preprocessed ECG data. The raw ECG signal was filtered as described in the TMSI<sup>6</sup> manual: First a high pass filter (8Hz) was applied to filter out large fluctuations in the signal. A 15ms second delay was added, together with a delta filter to let the low frequency parts of the signal disappear. To be independent of the direction of the QRS complex (due to morphology of the ECG), we took the absolute signal. Finally, a moving window averager (0.1sec) was added to get the envelope of the signal. This yielded a filtered signal with clear peaks. The raw and preprocessed ECG data will be made available.

We also calculated the heart rate and heart rate variability. Therefore, we processed the filtered data further in Matlab. First of all we applied a peak detection algorithm to the filtered signal. To determine the heart rate, the found

<sup>5</sup>We use a projection to the plane to distill only variance in one direction.

<sup>6</sup><http://www.tmsi.com>

peaks were counted per 1 minute time-frame. Then we calculated the distance between the found peaks (R-R). To determine the heart rate variability we took the root mean square of all these peak distances (RMSSD). Due to some remaining noise in the signal, the peak finding algorithm sometimes failed to accurately detect peaks. Therefore we excluded all 1-minute time frames in which more than one peak distance appeared unusual. We defined an unusual peak distance as a distance larger than 1.2 seconds where probably a peak was missed (or otherwise the HR would be below 50bpm) or a distance smaller than 0.5 seconds where probably an extra peak was detected (or otherwise the HR would be over 120bpm). The resulting heart rate and heart rate variability are contained in the feature dataset.

Moreover, we provide raw skin conductance data. We also calculated the average skin conductance level by averaging the raw signal per minute, which is contained in the feature dataset.

### 3.5 Example Analyses

In this section we present some research that was done based on our dataset, as an example of its use.<sup>7</sup>

**Work stress.** To find relations between the measured concepts, we performed a correlation analysis on the questionnaire data. We found that perceived stress is moderately related to high task load in terms of mental demand, temporal demand and frustration. Moreover, stress is related to emotion in terms of negative valence and high arousal. For more details on these results, see Koldijk, Sappelli, Neerincx, *et al.* (2013).

To investigate the effect of our stressors on the participants' subjective experience, we compared the questionnaire ratings of the neutral baseline condition with the time pressure and email interruption conditions (see Table 6.2). Under the stressor time pressure, participants experienced significantly higher temporal demand and higher arousal. The stressor email interruptions yielded reports of more mental effort, more positive valence and more dominance. We found that perceived stress did not differ significantly between the stressor and neutral conditions. Stress might be a too complex concept to measure in a short-termed work task. For more details on our results, see Koldijk, Sappelli, Neerincx, *et al.* (2013). These analyses show the potential of using the dataset for research on the effect of work stressors on experience and behavior.

**User modeling.** We are aiming to develop algorithms that can estimate the level of workload and stress that a knowledge worker is experiencing from sen-

---

<sup>7</sup>In Chapters 5 and 6 we provide more elaborate analyses.

Table 3.3: Subjective experience data (one rating per block). Average values for the Neutral, Interruption and Time pressure condition in the last 3 columns.

Type	Feature	Description	N	I	T
TaskLoad (NASA-TLX)	MentalDemand (0: low - 10: high)	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)?	4.9	5.4	4.9
	PhysicalDemand (0: low - 10: high)	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)?	1.9	2.3	2.7
	TemporalDemand (0: low - 10: high)	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred?	5.7	5.9	7.1
	Effort (0: low - 10: high)	How hard did you have to work (mentally and physically) to accomplish your level of performance?	5.2	5.9	6.1
	Performance (0: poor - 10: good)	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?	4.8	6.1	6.0
	Frustration (0: low - 10: high)	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	3.5	3.6	3.5
Mental Effort (RSME)	MentalEffort (0: no - 10: extreme effort)	How high was the mental effort for the tasks you just finished?	5.5	6.5	6.3
Emotion (SAM)	Valence (1 - 9)	How do you feel at this moment? (unhappy-happy)	4.8	5.7	5.3
	Arousal (1 - 9)	How do you feel at this moment? (calm-excited)	3.3	3.9	4.6
	Dominance (1 - 9)	How do you feel at this moment? (submissive-dominant)	5.2	6.2	5.9
Stress (VAS)	Stress (0: not - 10: very stressed)	How stressed do you feel?	2.9	3.2	3.8

sor data, in order to unobtrusively model their mental state. To investigate whether the stressors affected the participants' behavior, we compared computer interactions and facial expressions in the two stressor conditions with the neutral baseline condition. Under time pressure we see significantly more key strokes than in the neutral condition, and under interruptions we see more application changes and left clicks. So both stressors create typical behavioral patterns.

Explorative correlation analysis on questionnaire and facial expression data showed that moderate correlations were found for mental effort with several facial features. When working in a condition with higher mental effort, participants looked more disgusted and sad and they showed more activation in the facial action units LidTightener, UpperLip-Raiser, BrowLowerer and CheekRaiser. So mental effort might be estimated based upon video information. For more details on our results, see Koldijk, Sappelli, Neerincx, *et al.* (2013). These analyses show the potential of using the dataset for research on user modeling.

**Context recognition and Information support.** The tasks of writing reports and preparing presentations enabled us to use the dataset in the fields of context recognition and search behavior as well. For this purpose, the raw events were aggregated in event blocks, which were labeled for their task content. Details can be found in Sappelli, Verberne, *et al.* (2014), where also some initial analyses of the search behavior in the data are described.

## 3.6 Discussion

To our knowledge, our dataset is the first in which a set of unobtrusive sensors from different research fields was used to collect data in a realistic office context, while stressors were manipulated.

In collecting and preprocessing the data, we encountered a number of challenges. First, simulating a realistic work setting and inducing stress was challenging. We do think that we succeeded in simulating a realistic work setting. Some participants noted afterwards, that although they knew that the emails were fake, they felt responsible to reply. We also think we were able to manipulate the working conditions with stressors: the questionnaire data showed that participants' experience of task load, mental effort and emotion changed in the stressor conditions. Due to our experimental design of always starting with the neutral baseline condition, subjects might have experienced order or fatigue effects. We do, however, think that the relaxation phases helped participants to start each condition in a well-rested state. We did not find significant effects of our working conditions on perceived stress. Real-world stress might be complex, involving worries or things outside work and stress building up over days. Therefore, a limitation of this dataset is that only short term effects of stres-



sors can be investigated. For longitudinal research on temporal (stress) patterns over days or weeks, we are currently recording the presented sensor suite in a real-world office.

A second challenge was synchronization of all data. Different sensors were recorded via different computers. We synced computer clocks, and most sensors made exact starting timestamps upon hitting the record button. Nevertheless, we cannot guarantee second-precise synchronization among modalities (especially the uEye camera start times may be somewhat unprecise).

Finally, using different sorts of sensors requires multidisciplinary expertise, like knowledge (and software) for processing physiological, image or Kinect data. Our contribution is to provide a dataset that not only contains raw data, but also preprocessed and aggregated data, which makes it easier for other researchers to use the data.

The strength of this dataset is its richness in terms of modalities and its size in terms of the amount of data per participant. Although limited in size of participants (25), initial experiments have shown that the dataset has sufficient power to detect significant differences. With the presented new dataset, automatic inference of a rich set of context information of a user in the office can be studied. It can be used to develop context aware systems to support knowledge workers during their work. Moreover, it provides ample resources for stress and work style related studies.

### 3.7 Conclusion

We identified the need of a rich dataset and its desired characteristics. In this chapter we described how we collected such a new dataset that overcomes drawbacks common in related work: We used a realistic office setting while stressors were manipulated systematically. We used a varied set of sensors: computer logging, video, Kinect 3D and body sensors. We preprocessed the data and extracted features per minute. The resulting dataset SWELL-KW will be shared with the scientific community. We presented a selection of research questions that could be answered with this dataset. As demonstrated, analyses of the data can yield insights in the effects of stressors at work, or on the relation between subjective ratings and the sensor data. The presented new affective and behavioral dataset is a valuable contribution to research fields like work psychology, user modeling and context aware systems.

The dataset has been curated and is archived sustainably by the DANS archive of KNAW at <http://persistent-identifier.nl/?identifier=urn:nbn:nl:ui:13-kwrv-3e>. More information on the dataset can be found at <http://cs.ru.nl/~skoldijk/SWELL-KW/Dataset.html>.

## Real-time task recognition based on knowledge workers' computer activities

*In this chapter, we present work on automatic task recognition.<sup>1</sup> Computer interaction data of knowledge workers was logged during their work. For each user different classifiers were trained and compared on their performance on recognizing 12 specified tasks. We found that after only a few hours of training data reasonable classification accuracy can be achieved. There was not one classifier that suited all users best. We conclude that task recognition based on knowledge workers' computer activities is feasible with little training, although personalization is an important issue.*

---

This chapter is based on Koldijk, van Staalduinen, Neerincx, *et al.* (2012). "Real-time task recognition based on knowledge workers' computer activities". Published in: Proceedings of the European Conference on Cognitive Ergonomics (ECCE 2012).

<sup>1</sup>The research on task recognition was performed during my Master internship. As the topic was relevant for my PhD thesis we decided to put effort in distilling the main findings and making a publication for the European Conference on Cognitive Ergonomics.

## 4.1 Introduction

Nowadays, many people spend their working days at a computer, coordinating different activities in several projects to create information products. We refer to these people as knowledge workers. Typically, they have to self-manage their work to accomplish all their tasks. Their course of action is not always self-planned but also determined by external causes, like phone calls, mails, information requests, other persons or appointments (Czerwinski, Horvitz, and Wilhite, 2004b), which easily results in a fragmented way of working. So, a good overview of tasks is important for them, but rather difficult to maintain. The goal of our research is to support knowledge workers with tools. This chapter aims at automatic task recognition to provide overviews of tasks performed.

Knowledge workers rely on software for communication, information gathering, document creation and work planning, so a vast collection of digital traces is left behind on their computer. These are available in the form of mouse motion, click events, key presses and active window changes. We use these traces to automatically infer what task a user is currently performing. In this way we automatically create a real-time overview of tasks for the user in an unobtrusive way.

As research has shown, more awareness of one's own working process can have beneficial effects on the on-task behaviour and adherence to scheduled activities (Richman *et al.*, 1988). A study by Johnson and White (1971) showed that mere self-observation caused a positive change in behaviour. By being able to easily look back at their behaviour, knowledge workers might get a better grip on their work style and improve it. Cognitive load and stress might be decreased.

Some systems that provide overviews of computer activity exist (e.g. Slife<sup>2</sup>, RescueTime<sup>3</sup>), but they present low-level data in the form of time spent per application and websites browsed. They require the user to interpret for which task a specific program or website was used. In our research, minimal effort should be required from the user. So we aim at automatic recognition of tasks based on computer activities. We use not only application information, but also typical patterns of behaviour that originate from mouse and keyboard.

In the field of activity recognition, various activities are automatically recognized, for example activities in an adventure game (Albrecht *et al.*, 1997) or computer activities, such as filling in a form or planning a meeting (Rath, Devaurs, and Lindstaedt, 2009). These activities have rather clear structures, involving predefined steps (see Natarajan *et al.* (2008)). Therefore, often model-based clas-

---

<sup>2</sup><http://www.slifeweb.com>

<sup>3</sup><http://www.rescuetime.com>

sification is applied, with logical models assuming a plan library (e.g. Goldman, Geib, and C. A. Miller (1999)) or Markov models, modelling the sequence of actions in time (e.g. Albrecht *et al.* (1997)). Moreover, most models are applied to simple problems in a controlled environment. Our models will be tested in a field study. The recognition of knowledge workers' tasks on the basis of computer activities is a new domain with different characteristics, where tasks are less structured and task sequences are more spontaneous. Whether task recognition in this domain is feasible is thus a challenging research question.

This chapter is organized as follows. First, we describe our framework for task recognition, then we explain how we evaluated this framework in a field study. Thereafter, our analyses and results are presented, followed by a discussion and conclusions.

## 4.2 Task recognition framework

To recognize knowledge workers' tasks automatically, a framework is necessary that specifies the mapping from low level computer interaction data to performed tasks. The following components are required to realize this framework:

1. A set of task labels that users intuitively use.
2. A number of useful features obtained from computer interaction data.
3. Different classifiers that map low level activity features to the defined task labels.

These components are described in the next three subsections.

### 4.2.1 Task Labels

To obtain more knowledge about tasks that knowledge workers typically perform, and which task labels they intuitively use, we developed a questionnaire. In total 47 employees from TNO (Netherlands Organization for Applied Scientific Research) with various backgrounds and different functions completed this online questionnaire. The answers to the questions 'What tasks do you perform and how do you use your computer to realize this task?' and 'Describe a typical working day' were manually grouped into sets of similar answers. Task categories that clearly arose from the data were email, meeting and planning. These were mentioned by nearly anyone. Depending on the specific role or expertise of the knowledge worker several project tasks were mentioned, such as searching for information, analysing data, making a presentation or writing a report.

Many people also listed phoning, traveling, using social media, coffee breaks, talking with colleagues, doing some private Internet browsing, or having lunch.

The appropriateness of our identified task labels was confirmed by several knowledge workers. We investigated automatic task recognition for those tasks that are performed using a computer: Read mail, Program, Write mail, Write report/ paper, Organize/ archive data, Search information, Plan , Read article/ text, Make presentation, Make overview, Create visualization, Analyse data.

We learned that knowledge workers do not intuitively think in terms of applications to categorize their activities. They have a specific purpose or task in mind, which often requires the use of several applications. The tasks are in focus and the applications used depend on these tasks. Important to note is that some applications, like PowerPoint, are used for different tasks. Therefore task recognition is not a simple one-to-one mapping between an application and a task. Users also switch between different applications while executing one task, which became clear from the descriptions of some respondents. Our recognition model should be robust to this behaviour.

#### **4.2.2 Features**

Automatic task recognition requires relevant features. In our research, computer interaction data is used, which should be automatically logged. From this raw data useful features should be extracted, such that the classifier can discriminate between tasks.

We used uLog (software developed by Noldus Information Technology ) to log mouse and keyboard actions, as well as the applications used. Thereafter, this raw data was processed to extract relevant features. All these features were calculated for a 5 minute time segment, which we assume to be long enough to average out fluctuations, but fine grained enough not to lose useful information. In this way we calculated for example how often the user clicked within the 5 minute segment, or how much of the time a certain application was in focus within this 5 minute segment.

Mouse features include

- the number of clicks and scrolls within the time frame.

Keyboard features include

- the amount of characters and special keys typed,
- the number of spaces and backspaces.

Application features include

- the application that was mainly in focus during the five minute time frame,
- features for typical applications like Word or Outlook, which indicate what percentage of time these applications were in focus.

Other features used are

- the number of different applications used within the time frame,
- the number of switches between applications,
- the time of the day.

### 4.2.3 Classifiers

For mapping simple features to higher level tasks, a classifier is used. All features determined for one time segment are provided to a classifier, which assigns a task label to this time segment. As knowledge workers' tasks do not have a clear predefined structure, which could be modelled, we chose to use several common and rather simple data-based classifiers: KStar, Naive Bayes, Decision Tree, and Multilayered Perceptron.

For all classifiers we used Weka (Hall *et al.*, 2009) with default settings. To investigate which of the different classification principles is most suitable in our domain, we compared the performance and learning curves of these classifiers.

The reason to use a single time segment for classification is that it simplifies the model, which yields fast task recognition and requires a small number of parameters to be estimated. This seemed a good starting point to us. This model is easier to train, than more complex temporal models, where the label of a segment is also determined by information from previous time segments. Moreover, training a temporal model requires more ground truth labels than our model, and in a real-world setting such a large labelled dataset is difficult to acquire.

## 4.3 Approach for framework evaluation

To evaluate our task recognition framework, we performed a field study in which data was collected from knowledge workers who were performing their daily job. These workers regularly annotated which task they were performing. This annotated data set was then used for several analyses. We aimed to investigate how good our framework is, in terms of classification performance and learning speed, for recognizing tasks performed by different knowledge workers.

We now explain the tool to collect annotated data in a user friendly way and then describe the method of our user study.

### 4.3.1 Tool For Collecting Annotated Data

For our study, the participants had to annotate their activities with task labels while working at the computer. A simple pop up reminding them to indicate which task they were currently performing was perceived as very annoying. Therefore, we created a more user friendly data annotation tool, which makes the labelling easier by suggesting task labels to the user. Classifiers were trained on the initially collected dataset of our pilot study. These are then used to automatically classify the previous five minutes of user activity. The recognized task label of one of the classifier types is then presented to the user in a small pop-up. The user can look back at the suggested task labels of the previous hour and confirm or correct them (see Figure 4.1). This approach makes it easy to check or correct activity labels whenever the user wished to. After one hour of new data the classifier is retrained to optimally predict suitable task labels for this user.

Besides making labelling of activities easier, we added two types of visualizations to make the use of the program more interesting for the participants. The first visualization depicts the performed tasks as a pie chart (see Figure 4.2). This gives the knowledge workers the possibility to look back and see which kind of tasks they were mainly performing over the days. The second visualization shows the activities of the knowledge workers as a Gantt chart (see Figure 4.3). In this visualization they can easily see the course of activities over the day. Our idea was that presenting users these visualizations gives them insights in their way of working and makes it more important to them to correctly label their activities.

### 4.3.2 Method

The exact method we followed for collecting annotated user data is described in this section.

**Participants** Eleven knowledge workers employed at TNO volunteered to participate in our two week data collection period (10 male, 1 female). All participants typically spent most of their working day at the computer and carried out a diverse set of typical knowledge worker tasks.

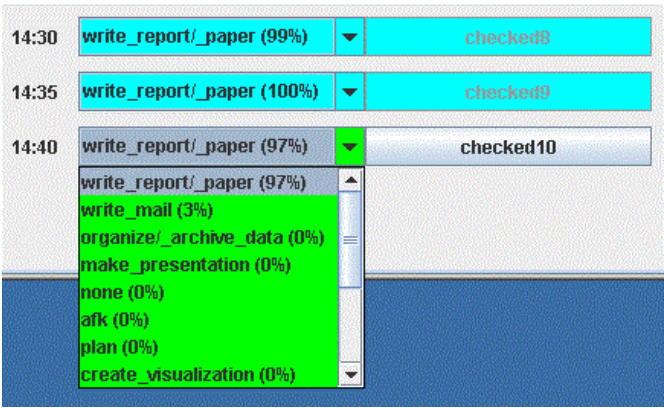


Figure 4.1: View to check or correct the automatic labelling.

**Materials** The participants worked at their regular work place on their own Windows desktop computer with mouse and keyboard. The logging tool uLog was installed on the machines to capture mouse, keyboard and application activity. The logging files were read out by a Java program and stored in a triple store database (Jena) on a server for further access. Another Java program was used to fetch the current activity data from the database (using SPARQL) and apply various classifiers from the Weka machine learning toolkit in order to suggest a task label to the user.

**Procedure** First of all the required software was installed on the participants' computers and its usage was explained shortly. The knowledge workers were instructed to start up the software at the beginning of the day and work as usual. During their work, the data capturing programs ran without attracting attention. Every five minutes, the recognition program analysed the user's activity data and suggested a task label to the user in a small pop-up window. All participants used this same setup. They were told to regularly check the suggested task labels and correct them when necessary, either immediately after the pop-up or within one hour via the dashboard view (see Figure 4.1). It was explained to the participants that they could access some simple visualizations of the activities of the days, which were automatically made, via the dashboard whenever they wished to.



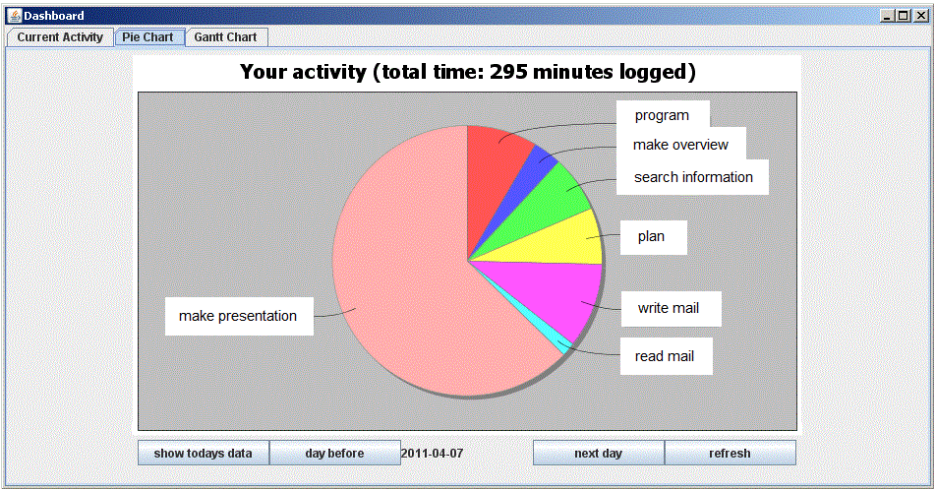


Figure 4.2: Dashboard with Pie chart visualization showing amount of spent time per task.

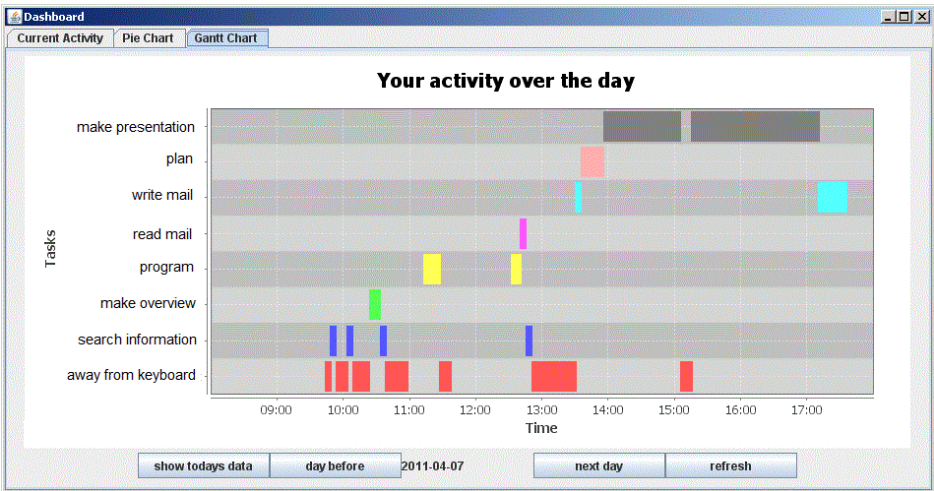


Figure 4.3: Dashboard with Gantt chart visualization presenting tasks performed during the day.

## 4.4 Analyses and results

The annotated data sets resulting from our field study were used for several analyses. In this section we present the analyses performed and the results obtained, beginning with a check on our chosen task labels and features. In the next subsection, the comparison of different classifiers will be described. Finally specific analyses regarding individual differences between users are presented. For results in full detail see Koldijk (2011).

The data collection phase resulted in eleven datasets, one for each participant. For a reliable ground truth only data with labels explicitly checked by the user were used in our analyses. In Figure 4.4 the amount of checked labels per user can be seen. As user J and B checked too few labels, their data was excluded from further analyses.

User	A	C	K	I	E	G	H	D	F	J	B
# labels	522	156	144	108	72	42	36	36	30	6	3
in hours	43.5	13	12	9	6	3.5	3	3	2.5	0.5	0.25

Figure 4.4: Dataset - amount of checked labels per user. (Users ordered on amount of data, users J and B were excluded from further analysis because of too little data.)

### 4.4.1 Task Labels and Features

First of all, we tested whether the defined task labels and the chosen set of features were suitable. Only the main insights are presented here (for more details see Koldijk (2011)). Regarding the task labels we considered confusion matrixes. In general, our task labels seemed appropriate. Typical confusions of tasks were mainly due to some tasks involving other tasks as subtasks (e.g. searching information being part of writing a document). Regarding our chosen features, we analysed their information gain. All our features turned out to be useful. Information about applications turned out to be a good feature among users, whereas mouse and keyboard activity as well as work style (e.g. switching behaviour) are good features on a per user basis.

### 4.4.2 Comparison of Classifiers

Next, we compared the selected classifiers in terms of performance and leaning speed. Details about the analyses and results are presented in the following two subsections.

**Performance** We used the Weka machine learning toolkit to train and test several classifiers, in order to answer the question which classifier is best in recognizing tasks. The performance of the classifiers was measured as percentage correctly classified instances. For performance evaluation we applied 10 fold cross-validation. To make the estimate more reliable we ran this whole process ten times and averaged the results over the runs.

Labelling each segment simply as the majority task with Weka's ZeroR classifier yielded us a baseline accuracy. We compared the performance of the following classifiers: KStar, Decision Tree, Naive Bayes and Multilayered Perceptron. All labelled data of one user at a time was used to train and test a classifier. This was repeated with all nine users' data sets.

As you can see in Figure 4.5, for each user all tested classifiers performed better than baseline (which was given by ZeroR). It differed per user which classifier achieved the best performance. For example, you see that the Perceptron was clearly best for user A with a final classification accuracy of about 70%, whereas for user I Naive Bayes gave best results with 80% accuracy. For user E KStar slightly won with 75% accuracy.

From our analysis we can conclude that the classification accuracy is reasonably high in this office setting, but it is impossible to say which of the classifiers generally achieves the best performance. The different classifiers use very different principles to discriminate between tasks. There is thus not one principle that clearly works best in this domain. It might depend on the specific work style or characteristics of the user which method is most suitable. We analyse the differences between users in more detail in the section on individual differences.

**Learning Curves** As a next step we investigated which classifier is fastest in learning to classify tasks. We simulated the growths of the data set in order to analyse the learning process of the classifiers. The user's complete data set was first of all split into 10 folds, one of these folds held apart for testing. From the remaining folds data was randomly sampled creating increasingly large training portions. The first training portion contained 3 sampled data instances, the next 6, 9 and 12 instances. From then on the training portion size grew with 6 instances (= half an hour of data). Every classifier was then trained on each of these training portions, always using the fixed test sets to evaluate their performance. We plotted the classifier performances for different data set sizes as learning curves (values again averaged over 10 test folds and 10 runs).

Figure 4.5 plots the learning curves per user. It shows that, in general, the performance of the classifiers was at 80% of its maximum after only about 30 instances, which is only 2.5 hours of training data. The particular form of the learning curves differed per user. For user K, all classifiers learned slowly,

whereas for user E they all learned quickly. For user I, there was a great difference between learning curves, with Naive Bayes quickly achieving a high performance and KStar performing badly, whereas for user K, all curves were mingled up, showing no clear winner in terms of learning speed.

From this analysis we can conclude that the classification is in general learned quickly in this setting, but it is impossible to say which of the classifiers generally learns quickest. Again, specific characteristics of the users seem to influence how fast a model is learned and which classifier is most suitable.

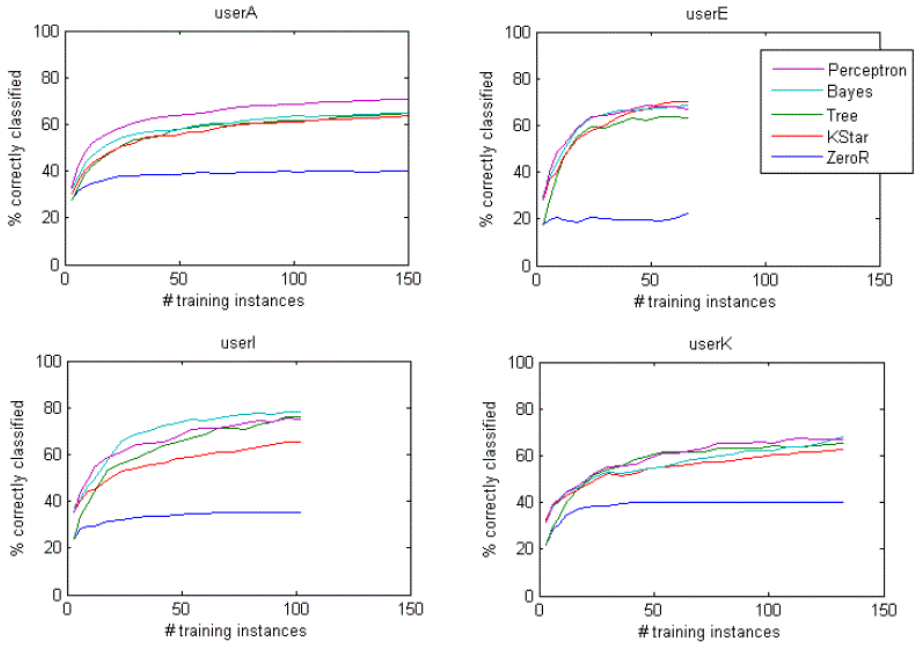


Figure 4.5: Learning curves for the different classifiers, for some selected users. Note: ZeroR provides a baseline.

#### 4.4.3 Individual Differences

We saw great variance in both final performance of the classifiers and their learning speed between users. This poses the following questions:

- Where do these performance differences come from, i.e. how do the users differ?

- Given these individual differences, how does a trained model perform on a new user?

**Differences Between Users** A first aspect we considered are the differences in the users' tasks. The distribution of tasks that the knowledge workers performed during the data collection period was analysed. Our results show that different users performed a different task mix. Some task combinations may be better distinguishable than others, so this can explain differences in classification performance.

A second aspect we considered is the typical pattern of behaviour of the users. Therefore we analysed the distribution of clicks, typing or other features per user and task. It turned out that even when users were performing the same task their behaviour differed (Koldijk, van Staaldin, Raaijmakers, *et al.*, 2011). For example user G typed extraordinary many characters when writing a report and in general clicked more often than other users. Statistical analysis in form of a 12 (tasks)  $\times$  9 (users) MANOVA with all features as dependent variables showed a significant effect of task and user on almost all features. This means not only the task, but also the specific users are distinguishable on basis of the measured behaviour. These results hint at different users having a different way of working. They might for example differ in work style, for example thinking a lot and typing a sentence in one go versus quickly typing and retyping things. Or they might differ in mouse use, for example using mainly the keyboard to navigate versus using the mouse to point and click. These individual characteristics also make task recognition more or less easy to learn for various classifiers. From these analyses we can conclude that the task mix of the users and their typical behaviour is very individual. This explains why there is no 'one classifier suits all' solution.

**Generalizability of the Classifiers** Analyses thus far indicate that task recognition is very personal. It is thus the question whether a classifier can be trained on a set of user data and effectively be used to classify a new user's behaviour.

To answer this question we first trained a classifier on the data of user A. We used this trained classifier to classify the test sets of all other users. Our results show that although the trained classifier worked fine on user A's test set it reached a performance of only 20% on average on other users' test sets. We can conclude from this that a classifier trained on one user does not work on other users' data.

Then we tested whether a classifier could become more robust in classifying a new user when it was trained on a mix of several users' data. The idea was that the classifier would not model specific details of one user, but pick up

general patterns common among users. We created training sets by sampling 30 instances per user of all but one user and trained classifiers upon these data sets. Then we tested its performance on the left out user's data to test the generalizability of the model. It turned out that the average classification performance was only 20 to 30% in this setting. This is better compared to training on one user's data, but far from satisfactory. We can conclude from this analysis that also a classifier trained on a mix of users' data does not generalize well to new users.

## 4.5 Discussion

Our research showed that task recognition in the domain of real-world knowledge worker activities is possible, but there is no clear recommendation to which type of classifier to use based on classification performance and learning speed. No classifier consistently worked best for all users. So, one might wish to consider other criteria to select the most suitable classifier. In the final application classification should be performed efficiently, without taking too much processing capacity. This makes KStar less suited, as classifying new instances can take long, because the dataset grows. Furthermore, the classifier needs to be regularly retrained in order to keep optimally adapted to the current behaviour of the user. From this perspective, the Perceptron approach seems less suited, because training on new data takes very long. Consequently a Decision Tree or Naive Bayes approach seem most suitable for task recognition in practice.

Furthermore, our research revealed that recognizing tasks on basis of computer activity is personal. Users differ in terms of the tasks they perform and how predictable or difficult their task mix is. Moreover, different users seem to have their own individual way of working. Besides the factors analysed here, other factors might be of influence too. Users might for example have different interpretations of what makes up a specific task and in how precisely they label their activities. Within one user, however, there is a general structure which makes task recognition possible. In general, we can state that a classifier can best be trained for one particular user. When the tool is applied to a new user, we face the so-called cold start problem. This problem can be solved by asking the user what he or she is doing at several moments during one week, thus collecting a representative set of annotated data for this user. As little as 2.5 hours (30 instances) of representative training examples is enough to train a good model. After this week the tool could start to recognize this user's tasks.

During our research we also gained some practical insights. First, some users reported that it was difficult to remember what exactly they had been doing. Some participants noted that the mere fact that they labelled their data made

them more aware of the tasks they were performing and some mentioned that this made them work more eagerly. So the data collection procedure, although designed to be unobtrusive, might have had some influence on the way of working. Another observation regarding data annotation was that the users were curious and interested in whether the tool would come up with correct labels, especially in the beginning, which motivated them to regularly check the labels. This curiosity and interest could be further exploited, making the annotation and the tool in general fun to use and game-like.

Furthermore, we observed in our experiment that users often think in terms of broader goals, not in terms of the specific methods used. This is in line with the differentiation that Heinze (2003) made (in Tahboub (2006)), describing an intentional level and an activity level. One might regard the more detailed description that the task recognizer comes up with as describing the specific activities performed, including all subtasks. Users agreed that they have actually performed these subtasks, but they themselves describe the tasks they performed during a day at a less detailed level, labelling only their intended main tasks. To capture this hierarchy of task labels one could take a series of subtasks over time to label the sequence with the intended main task label. Temporal models such as Markov models or conditional random fields could be considered for modelling these sequences, like is done in related research (e.g. Natarajan *et al.* (2008)). In this way, knowledge of tasks in general could be used to improve the classification, e.g. the fact that information seeking is often a subtask for another main task. One might also wish to use more flexible or overlapping time frames in order to find the exact beginning of new tasks. Nevertheless we see no need to make an overly complex model when with a simple model acceptable accuracy can be reached.

Enabling automatic task recognition is a first step of the SWELL project<sup>4</sup>. With a broader view on the context and mental state of the knowledge worker, we aim to provide optimal support to improve well-being at work. Clearly not all work of a knowledge worker can be captured on basis of computer activity, e.g. time spent in meetings, phone calls, talks with colleagues or reading printed documents. To get a more complete view, we intend to make use of other sources of information. For situations when the user is not active on the computer, we can use information from the user's calendar to fill in gaps. We can also use a camera and microphone to get more information about the user's current situation, like talking to colleagues. Moreover the mobile phone can be a very valuable source of information with call logs, and built in accelerometers and GPS to infer movement and location of the user. We also intend to infer the

---

<sup>4</sup><http://www.commit-nl.nl/Smartreasoningsystemsforwell-beingatworkandathome>



content the user is working on from documents on the computer. Besides that, estimating the mental state of the user is of interest, like the workload and stress level (Koldijk, Neerincx, and Kraaij, 2012). With this information, optimal support and coaching could be provided.

## 4.6 Conclusions

In this chapter we have presented task recognition based on computer activities in a real-life setting. Our research has shown that task recognition on the basis of PC activity is challenging but feasible.

First, task recognition involves more than a simple one-to-one mapping between an application and a task. This is due to interleaved activities, switches to subtasks and a mix of applications used that determine the task performed by the user.

Second, task recognition is very personal. Different users have different work styles and task mixes. Nevertheless, we saw that on an individual basis, the classifiers we used learn to recognize tasks quite fast, yielding a performance up to 80% which is reasonable high, considering the 12 possible task labels that are used.

Third, unlike other research, in which clearly structured tasks were modelled (see e.g. Natarajan *et al.* (2008)), our research has shown that task recognition also works for less structured tasks and more spontaneous activity, since our results were obtained using realistic data.

Fourth, comparison of several classifiers revealed that there is not one classifier that clearly works best in this domain.

Finally, since different users show different patterns of behaviour when performing a task, the classification model should be trained for each specific user to yield optimal task recognition. We concluded that no more than 2.5 hours (30 instances) of representative training examples is required to train a good model.





## Visual analytics of work behavior data - Insights on individual differences

*In this chapter, we describe how we applied visual analytics to the SWELL knowledge work dataset, containing information on facial expressions, postures, computer interactions, physiology and subjective experience. The challenge is to interpret this multi-modal low level sensor data. In this work, we alternate between automatic analysis procedures and data visualization. Our aim is twofold: 1) to research the relations of various sensor features with (stress related) mental states, and 2) to develop suitable visualization methods for insight into a large amount of behavioral data. Here we mainly focus on facial expressions, in Chapter 6 we analyze all modalities in more detail. Our most important insight is that people differ a lot in their (stress related) work behavior, which has to be taken into account in the analyses and visualizations.*

---

This chapter is based on: Koldijk, Bernard, *et al.* (2015). "Visual Analytics of Work Behavior Data - Insights on Individual Differences". Published in: Proceedings of the EG/VGTC Conference on Visualization (EuroVis 2015).

## 5.1 Introduction

Stress at work is a serious problem, in the worst case leading to burn-out. The goal of the SWELL project (<http://www.swell-project.net>) is to help employees to detect patterns in work behavior with potential danger to well-being in time. Trends like ‘quantified self’ show a potential of collecting personal sensor data (e.g. heart rate, activity patterns) for health improvement. Vast amounts of data from different sensors can be recorded, yielding a multi-modal, time oriented, and multivariate data set. Interpreting this data in a meaningful way is challenging. In this chapter we describe how we applied visual analytics to a work behavior dataset. We alternated between data analysis to find structures in the data and visualizations to gain insights. This exploratory analysis was aimed at finding relations between work stress and behavior that can be measured with sensors.

**Data:** We used the multi-modal SWELL-KW dataset (Koldijk, Sappelli, Verberne, *et al.*, 2014), which was collected in an experiment in which 25 people performed typical knowledge work (writing reports, making presentations, reading e-mail, and searching for information) for about 3 hours. We manipulated working conditions. In the neutral condition, participants were instructed to work as they would usually do. In one stressor condition, the participants got email interruptions and in the other stressor condition, participants worked under time pressure. As ground truth, questionnaire ratings of the participants were accessed after each condition for task load, mental effort, emotion and perceived stress. Work behavior data was recorded with various sensors: computer logging, camera, Kinect 3D sensor, and physiological body sensors. For the analyses presented here we used the preprocessed feature dataset, which contained averages per minute for several extracted features regarding: computer interaction, facial expressions, body postures, heart rate (variability) and skin conductance. We also had topic labels for what the participants were working on (Sappelli, Verberne, *et al.*, 2014). The SWELL-KW dataset can thus be described as a multi-modal, time-oriented, and multivariate dataset. The dataset consists of 3000 data entries of 25 different participants (on average 120 data points per participant) and has 150 columns containing different features.

**Research question:** How can sensor data be used to gain insight into work behavior, specifically related to stress at work?

**Contributions:** 1) Research into the relations of various sensor features with (stress related) mental states; 2) Focus on differences between users and how to cope with them in data processing and visualization; 3) Towards visualizations for a large amount of behavioral data recorded with sensors.

## 5.2 Related work

### 5.2.1 Affective computing

Regarding the relation between stress and sensor data, Sharma and Gedeon (2012) provide a compact survey. Often, **body sensors** are used to measure the physiological stress response directly (e.g. skin conductance (J. Bakker *et al.*, 2012), heart rate (Hogervorst, Brouwer, and W. K. Vos, 2013)). There also is potential in using outward characteristics, such as facial expressions, postures or computer interactions as indicators for the user's mental state. **Facial expressions** are currently mainly used for inferring emotions, but people might also show facial expressions indicative of mental states. Craig *et al.* (2008) looked at facial expressions while students worked with an online tutoring system. Association rule mining identified that frustration was associated with activity in facial action units 1, 2 (inner and outer brow raiser) and 14 (dimpler); confusion was associated with AU 4 (brow lowerer), 7 (lid tightener) and 12 (lip corner puller). Moreover, preliminary results by Dinges *et al.* (2005) suggest that high and low stressor situations could be discriminated based on facial activity in mouth and eyebrow regions. Regarding **postures**, Kapoor and Picard (2005) present research in which posture data was successfully used for estimating interest (vs. uninterest). Mental states are also being estimated from **computer interaction** data. Results by Vizer, Zhou, and Sears (2009) indicate that stress can produce changes in typing patterns. We think also general human computer behavior, such as task switching or browsing, might be indicative of mental states. In general, the affective computing community often uses (black-box) machine learning algorithms to classify sensor data into mental states (see Sharma and Gedeon (2012)). Often one model is learned over all users. This work is different, as we are interested in finding underlying behavioral patterns related to stress, for individual users. Visualizing behavioral patterns may give users more insight and actionable information than just a stress labeling.

### 5.2.2 Visual analytics

The structures of our included data set fit well to the survey of Kehrer and Hauser (2013) for the visualization and visual analysis of multi-faceted scientific data. Most relevant are the characterizations for multi-modal, multivariate, and (spatio-) temporal data. From a task-based perspective we borrow concepts from exploratory data analysis. In particular, most relevant classes of techniques are feature selection, visual comparison, and feature relation. In our approach we have to identify and **select features** relevant to mental states. Regarding human motion analysis, a visual-interactive exploratory search system was presented

in Bernard, Wilhelm, *et al.* (2013). The results of the data characterization phase give an indication on the complexity of the feature selection and (pre-) processing tasks for respective data sets. A profound overview of feature selection techniques in general is presented in the survey of Guyon and Elisseeff (2003). Well-known examples for the visual-interactive specification and selection of features are the Polaris system (Stolte and Hanrahan, 2000) and approaches presented by Doleisch, Gasser, and Hauser (2003) and May *et al.* (2011). A **visual comparison** helps to show differences in behavior related to stress. A taxonomy for information visualization tools that support comparison tasks is provided by Gleicher *et al.* (2011). They distinguish between juxtaposition, superposition, and explicit encoding. Visual comparison tasks can also be supported through compact representations of multidimensional data with glyph designs. For a recent state-of-the-art report on glyph-based visualization we refer to Borgo *et al.* (2013). An important step in our approach is the identification of **feature relations** within our multi-modal data set. Multiple linked views are an important class for supporting the visual-interactive exploration of relations. In addition, superposition-based overview visualizations can support users in revealing relations between multi-modal features. In the approaches of Bernard *et al.*, the users are guided towards interesting relations between clusterings of time series data and attached meta-data attributes (Bernard, Ruppert, M. Scherer, Schreck, *et al.*, 2012; Bernard, Ruppert, M. Scherer, Kohlhammer, *et al.*, 2012). In addition, visual-interactive approaches addressing relation tasks for mixed data sets are presented in Kosara, Bendix, and Hauser (2006), Alsallakh *et al.* (2012), and Bernard, Steiger, *et al.* (2014).

## 5.3 Data analysis and visual analytics

We carried out an iterative process by moving back and forth between data analysis and data visualization. In this process we gained a variety of insights in the nature of the dataset and its challenges, which we then aimed to address in the next iteration. Here we describe this process in detail.

### 5.3.1 General overview

We started with some general analyses on the dataset (Koldijk, Sappelli, Neerincx, *et al.*, 2013). A comparison of our working conditions (neutral, stressor email interruptions, stressor time pressure) with t-tests revealed that people showed differences in computer interactions under stressors. However, t-tests did not reveal an increase in experienced stress under stressors. As this result surprised us, we plotted the difference in stress for each participant in a

bar chart. This visualization revealed that for half of the participants perceived stress was higher under a stressor, whereas for a quarter of the participants stress was lower. When comparing conditions with t-tests, simple averages over all participants were taken, which kept such effects hidden. This brings us to our first main insight: Data visualization enabled us to view all 25 individual users and eased the identification of effects within subgroups. To be able to look into the data of individual participants, we implemented a visualization for the SWELL-KW dataset ([http://cs.ru.nl/~skoldijk/Visualization/ExperimentBrowser/Generic/Gantt\\_and\\_Numeric.html](http://cs.ru.nl/~skoldijk/Visualization/ExperimentBrowser/Generic/Gantt_and_Numeric.html)). Subjective experience data is displayed in relation to different sorts of data. For sensor data we use line charts to plot specific features over time. We applied machine learning techniques to the computer logging data to infer the current task (Koldijk, van Staaldin, Neerinx, *et al.*, 2012). The categorical variables are plotted as Gantt charts (see Fig. 5.1). This is a starting point for giving insight into working behavior. The question that arises is: *Are there interesting relations between subjective variables and features measured by the sensors?*

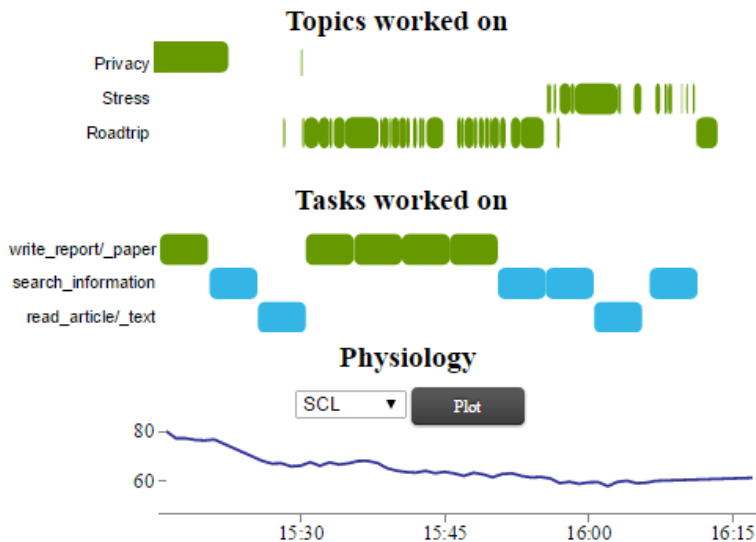


Figure 5.1: SWELL-KW dataset browser (extract). Topic and task worked on depicted (green: main task, blue: help task), together with the skin conductance level (measure of stress).

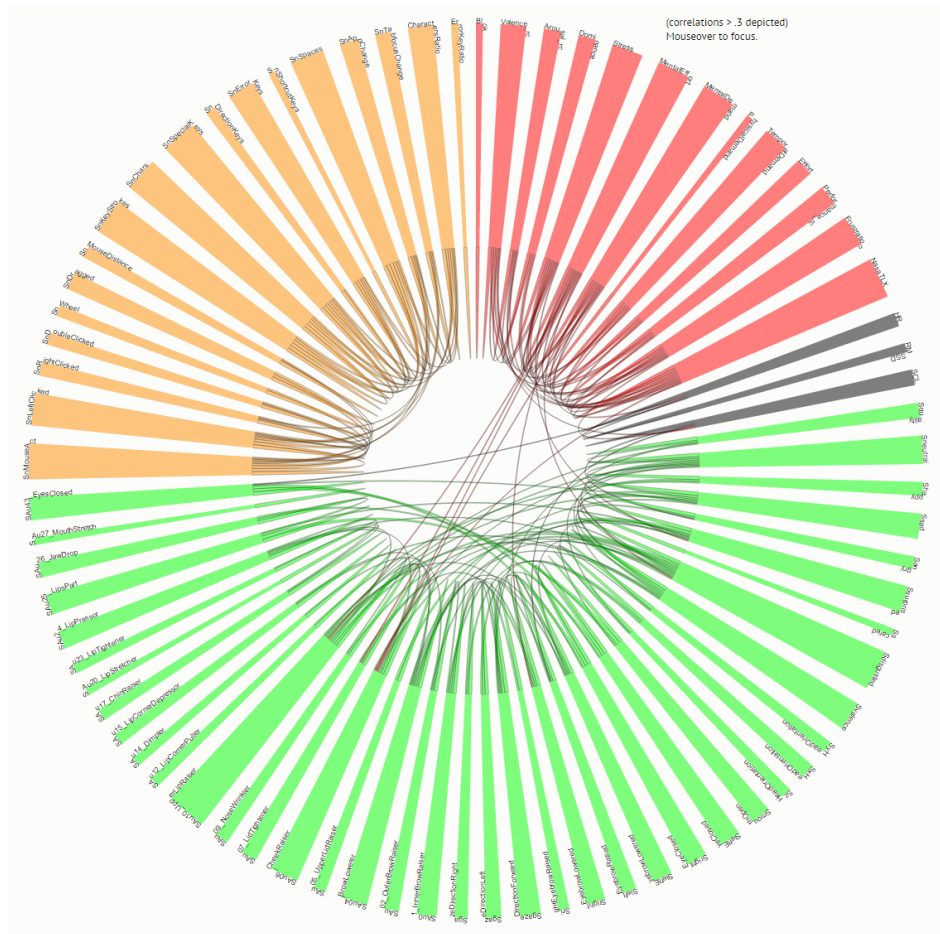


Figure 5.2: Interactive chord diagram of the correlations found in the SWELL-KW dataset. On hovering over a variable, all its connections become highlighted. The following groups of variables are depicted: questionnaire (red), physiology (gray), facial expressions (green) and computer logging (yellow) (posture features omitted for readability).

### 5.3.2 Relating subjective experience to sensor data

To investigate which subjective experience variable (stress, task load, mental effort, emotion; from questionnaires) is most closely related to measurable sensor data we performed a correlation analysis. For easier interpretability we decided to visualize the 150x150 correlation matrix as a chord diagram (Figure 5.2<sup>1</sup>). On the circle, several variables are depicted. When a correlation stronger than 0.3 was found in the data, a connection is drawn between two variables. In the image, many connections between variables of the same source (same color) can be identified. More interesting for us, however, are connections between subjective variables (red) and the sensors. The most promising relation seems to be the one between mental effort and several features resulting from facial expression analysis (green). The question now is: *How can we read someone's mental effort from facial activity?*

We used our dataset browser to plot the facial activity features most strongly correlated with mental effort, expecting to see clear trends, e.g. increase in lid tightening as indicator of increased mental effort. The results were disappointing. In general there is much fluctuation over time in the sensor data. Moreover, we see much variation between users: for some users specific action units are very active, whereas others show no activity in the specific facial region at all. The features with high correlations overall are not really insightful for individual participants. Some of the more 'extreme' users had a big influence in the correlation analysis. By using our visualization we were able to reveal individual differences, which need to be addressed. The question that arises is: *Can we group participants in their characteristics for further analysis?*

### 5.3.3 Typical user behavior groups

For the identification and comparison of different user groups we applied clustering techniques. Hierarchical clustering was used to reveal the amount of clusters (k) in the data and then k-means clustering was applied. We addressed each sensor separately and found that for each sensor the users were grouped differently. Clustering of computer activity data revealed that users differ in how they work: one group could be described as 'copy-pasters' (many special keys, a lot of mouse), the other as writers (many keystrokes). Clustering of facial activity data revealed that one group shows little facial activity, one group shows tight eyes and a loose mouth, whereas the other group shows wide eyes and a tight mouth. Clustering of body movement data revealed that one group

---

<sup>1</sup>see also <http://cs.ru.nl/~skoldijk/Visualization/ExperimentAnalysis/CircleCorrelation-noKinect.html>



sits rather still with their body and moves the arm a lot, another group moves the entire body, the last group just moves average. As users seem to differ in their general behavior, we normalized the data to make different users comparable: from each value, we subtracted the participant's average. Thus we are now focusing on difference scores, e.g. how much more than average someone is frowning. The question is: *Can we find strong features among all participants with this normalized dataset?*

### 5.3.4 Filtering the set of features

To find the most relevant features (from the available >150 features) to predict mental effort, we decided to apply (information gain based) feature selection to our normalized dataset. Again, we plotted the best features with our dataset browser. But even with the data normalized per user, the best features in general do not seem to give any insights for individual users. The change in behavior still is very individual. *We decided to focus on individual users and further explore their facial expressions.*<sup>2</sup>

### 5.3.5 Exploration of facial activity patterns

To investigate whether there are any meaningful patterns in the facial activity of individual users, we used the Acume behavioral data visualization toolkit (McDuff *et al.*, 2011). The tool generates heat maps from facial action unit data and facilitates the comparison of different users and different working conditions. A comparison of users showed large individual differences regarding which facial regions show activity in general. When comparing data within a user, however, differences in facial activity between working conditions become visible. Inspired by the Acume toolkit, we implemented a heat map visualization, which enables us to see patterns in different features in one glance (Figure 5.3). This is an improvement over separate line charts, but *to make the data better interpretable we think an avatar displaying the particular facial expressions would be an useful addition.*

### 5.3.6 Details on demand: Visualizing facial activity

To render facial expressions on an avatar (see Figure 5.4) we used the Hap-FACS tool (Amini and Lisetti, 2013). We added this visualization as a detail-on-demand concept to the heat map visualization<sup>3</sup>. In addition to seeing general

---

<sup>2</sup>In Chapter 6 we analyze all modalities in more detail.

<sup>3</sup>see <http://cs.ru.nl/~skoldijk/Visualization/ExperimentBrowser/heatmap/FacialActionUnits2/heatmap3.html>

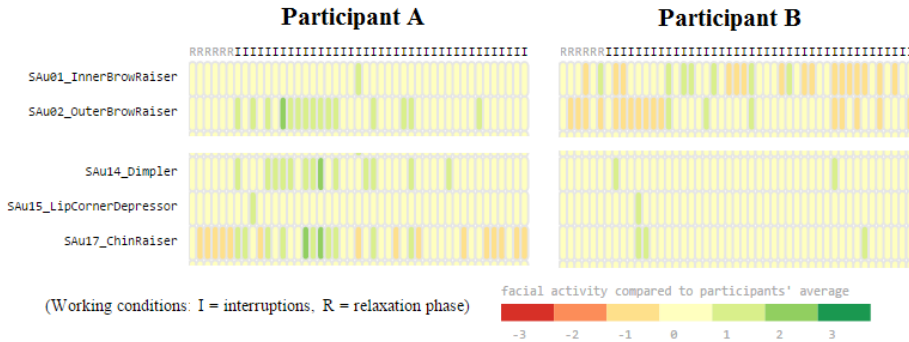


Figure 5.3: A heat map of facial activity data of 2 participants. Different facial regions show changes in activity.



Figure 5.4: HapFACS avatar displaying different facial expressions of one participant while working

patterns in the heat map, the user can hover over a point in time to get a representation of the actual facial expression. *Ideally we would want to add an indication of which facial expressions typically refer to high mental effort.*

### 5.3.7 Grouping typical facial expressions

To find typical facial expressions related to high mental effort, we applied a supervised Self-organizing Map (SOM). The SOM takes facial action unit data together with annotations on mental effort and produces a map in which the regions are ordered according to mental effort. Each cell of the SOM contains a typical facial expression, which was visually encoded with a Chernoff face metaphor (Fig. 5.5; alternatively the avatars could be used). As a result, we found several facial expressions typically associated with a high or low mental effort.

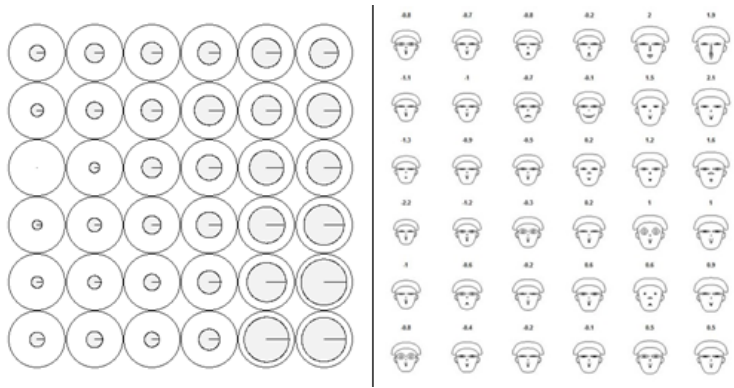


Figure 5.5: Left: Self organizing map, sorted on mental effort; Right: Facial expressions related to mental efforts.

### 5.4 Conclusion and future work

In an iterative approach, we used automatic data analysis procedures and visualization techniques in order to answer our research question: How can sensor data be used to gain insight into work behavior, specifically related to stress at work? We found that mental effort seems to be most closely related to facial expression features. Which specific facial action units are most informative differs per user. By clustering we were able to identify several user groups. Even after normalizing our data per user, individual differences remained. By means of a heat map we were able to visualize meaningful patterns in facial activity for an individual user. The visualization was made more insightful by rendering facial expressions on an avatar. Finally, we identified several facial expressions that are typically related to a low or high mental effort. We conclude that facial expressions may be a promising measurable outward characteristic that can be visualized to indicate mental state patterns during work.<sup>4</sup> The benefit of incorporating visual analytics to our problem, instead of a black box machine learning approach, was to gather a deeper understanding of the structures in our data and to gain insights from individual users' data.

Important lessons learned are: 1) There are many individual differences. People experience stressors differently, people differ in their usual work behavior and people show different outward behavior under stress; 2) Machine learning techniques that build one general model over all users seem not to make sense under these conditions. Models should be trained on individual users or

<sup>4</sup>In Chapter 6 we analyze the use of all modalities for estimating mental states.

groups of similar users. 3) A direct mapping from low level sensor data to subjective experience is hard. We rather suggest to first interpret low level data to a higher level e.g. raw computer interactions to tasks, facial action unit activity to meaningful facial expressions.

The presented results can serve as a baseline for a variety of future approaches. In future work, detailed analysis of specific users or user groups should be done, ideally with a dataset containing more labeled data per user. We now mainly focused on facial expressions. The combination of data from different sensors might be interesting to analyze and visualize. The presented analyses and visualizations should be integrated in an interactive visual analytics system. Finally, a user-study should be performed to evaluate whether the resulting visualizations can indeed help employees to detect alarming patterns in work behavior.



## Detecting work stress in offices by combining unobtrusive sensors

*The focus of this chapter is developing automatic classifiers to infer working conditions and stress related mental states from a multimodal set of sensor data (computer logging, facial expressions, posture and physiology; as described in Chapter 3). We address two methodological and applied machine learning challenges: 1) Detecting work stress using several (physically) unobtrusive sensors, and 2) Taking into account individual differences, as this was an important aspect revealed in Chapter 5. A comparison of several classification approaches showed that, for this dataset, neutral and stressful working conditions can be distinguished with 90% accuracy by means of SVM. Posture yields most valuable information, followed by facial expressions. Furthermore, we found that the subjective variable 'mental effort' can be better predicted from sensor data than e.g. 'perceived stress'. A comparison of several regression approaches showed that mental effort can be predicted best by a decision tree (correlation of 0.82). Facial expressions yield most valuable information, followed by posture. With respect to individual differences, we see that information on the particular participant improves accuracy. So, especially for estimating mental states it makes sense to address individual differences. When we train models on particular subgroups of similar users, (in almost all cases) a specialized model performs equally well or better than a generic model.*

---

This chapter is based on Koldijk, Kraaij & Neerincx. "Detecting work stress in offices by combining unobtrusive sensors". Submitted to: IEEE Transactions on Affective Computing.

## 6.1 Introduction

Employees often report the experience of stress at work, which can in the worst case lead to burn-out. Stress is a broad concept referring to psychological and biological processes during emotional and cognitive demanding situations. We follow a pragmatic approach and decompose stress in three factors that can be measured more precisely: (1) the task load, which poses demands on the worker, (2) the mental effort, which the worker needs to handle the task and (3) the emotional response that is raised, in terms of arousal and valence.

In the area of stress research, questionnaires are commonly used to get insight in the general working experiences (e.g. Zapf (1993)), but little is known on the immediate effects of stressors at work. Work in the area of affective computing investigates the possibility of inferring stress and emotion from sensor data (see e.g. Matthews, McDonald, and Trejo (2005)). To investigate the direct effect of different degrees of mental load, typically standardized tasks are used in a lab setting, such as remembering digits. These tasks are very simple and not representative of ‘real’ office work. Furthermore, work on user state modeling is often performed in a process control context, e.g. on naval ships (Neerincx, Kennedy, *et al.*, 2009) or in flight control. Only little work is done on user state modeling in an office context.

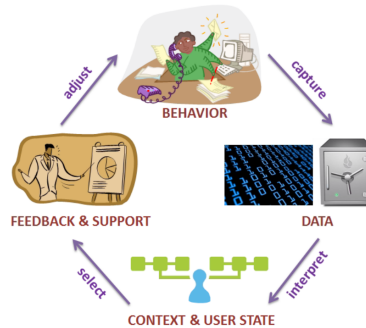


Figure 6.1: SWELL approach.

In the SWELL project we investigate how unobtrusive and easily available sensors can be used in offices, to detect stress and the context in which it appears in real-time (see Figure 6.1; Koldijk (2012)). Based upon this information, we aim to develop pervasive supporting technology that is optimally adapted to the current working context and mental state of the user. Knowledge workers can then directly act, gaining a more healthy work style and preventing stress

building up. Trends like ‘quantified self’ already show the potential of collecting personal sensor data (e.g. heart rate, activity patterns) for health improvement. Personal sensor data is relatively easy to collect nowadays, the challenge is making sense of this data.

The focus of this chapter is on developing automatic classifiers to infer working conditions and stress related mental states from a multimodal set of sensor data: computer logging, facial expressions, posture and physiology. We present related work in Section 6.2. The dataset that we use is presented in Section 6.3. We identified two methodological and applied machine learning challenges, on which we focus our work:

1. **Using several unobtrusive sensors to detect stress in office environments.** We found that state of the art research in stress inference often relies on sophisticated sensors (e.g. eye tracker, body sensors), and/or uses data collected in rather artificial settings. We see possibilities to build human state estimation techniques for use in office environments. We aim to combine information from multiple weak indicator variables based on physically unobtrusive measurements. We address the following **research questions**: Can we distinguish stressful from non-stressful working conditions, and can we estimate mental states of office workers by using several unobtrusive sensors? Which modeling approaches are most successful? Which modalities/ features provide the most useful information? This helps to configure a minimal sensor set-up for office settings. We address these questions in Section 6.4.
2. **Taking into account individual differences.** We found that, in affective computing, often one generic model is learned for all users. This may work for something universal, as the expression of emotions. However, in earlier work (Koldijk, van Staalduinen, Raaijmakers, *et al.*, 2011; Koldijk, Bernard, *et al.*, 2015), we found that people differ in their (work) behavior: typical behavior of users already differs per person. Moreover, the way in which people express mental effort or stress may differ. This highlights a need to build personalized models for particular users or user groups, instead of one general model. We address the following **research questions**: How important are individual differences? Can we improve performance by building personalized models for particular user groups? We address these questions in Section 6.5.

Finally, we present our Conclusions and Discussion in Section 6.6.



## 6.2 Related work

Here we present related work on affective computing, more particularly on using physiology, facial expressions, postures and computer interactions (or preferably a combination of modalities) to infer the user's mental state, e.g. in terms of stress. We describe the sensors that are used, in which context data is collected, which machine learning approaches are applied, and how individual differences are addressed.

**Physiology** Most often, body sensors are used to measure the physiological stress response directly. For a general overview of psycho-physiological sensor techniques we refer to Matthews, McDonald, and Trejo (2005). Most studies report on experimental environments, since this new technology is hardly deployed yet in work environments.

In research by Riera *et al.* (2012), for example, brain imaging (electroencephalography, EEG) and facial electromyography (EMG) data were collected. The authors show that these sensors can be used for monitoring emotion (valence and arousal) and stress. Although its great potential, we think deploying brain imaging in a daily office setting is not yet realistic.

Other common measurements in stress research are pupil diameter and heart rhythm (electrocardiogram, ECG). Mokhayeri, Akbarzadeh-T, and Toosizadeh (2011), for example, collected such data in context of the Stroop color-word test. They state that pupil diameter and ECG have great potential for stress detection.

The question that arises is: can we make an estimate of affective and mental states outside the lab? We see some potential for heart rhythm measurements (ECG) and skin conductance (GSR or EDA), with the rise of wearable sensors, which are becoming more and more integrated into devices as watches and bracelets. J. Bakker *et al.* (2012) e.g. measured skin conductance of 5 employees during working hours.

Setz *et al.* (2010) present work in which they use EDA measurements to distinguish cognitive load and stress. 32 participants solved arithmetic tasks on a computer, without (cognitive load condition) or with time pressure and social evaluation (stress condition). To address individual differences, data was also normalized per participant by using a baseline period. However, the non-relative features turned out to work better. Leave-one-person-out cross validation yielded an accuracy of 82% to distinguishing both conditions. The authors 'suggest the use of non-relative features combined with a linear classification method' (p.416).

Cinaz *et al.*, 2013 present research in which they used ECG. 3 calibration tasks in a laboratory setting were used to induce low, medium and high workload.

This was used to train models, which were then used on 1 hour data recorded during office work. Data was aggregated over 2 minutes. They find that linear discriminant analysis (LDA) performs best in predicting mental workload (classifying 6 of 7 participants correctly), followed by k-nearest neighbor (kNN) and support vector machines (SVM). Measuring physiological stress reactions by means of EDA and ECG in office settings may thus be feasible.

**Facial expressions** There also is potential in using behavioral cues, such as facial expressions, postures or computer interactions as indicators for the user's mental state. For an overview of machine understanding of human behavior we refer to the survey by Pantic *et al.* (2007). In related work, facial expressions are widely used for inferring emotions. The data are often recorded while emotions are induced in participants. The publicly available multimodal dataset described by Soleymani *et al.* (2012), for example, was collected in context of watching emotion inducing video clips and consists of: face videos, audio signals, eye gaze data and physiological signals (EEG, ECG, GSR, respiration amplitude, skin temperature). Although this dataset is very interesting, emotions in a daily computer work context are probably less intense than the valence or arousal experienced during watching a movie clip.

An interesting question is whether people show facial emotions during computer work, and whether their facial expressions are indicative of mental states. Preliminary results by Dinges *et al.* (2005) suggest that high and low stressor situations could be discriminated based on facial activity in mouth and eyebrow regions. They applied a Hidden Markov model. They state that their algorithm has 'potential to discriminate high- from low-stressor performance bouts in 75 - 88% of subjects'.

Moreover, Craig *et al.* (2008) looked at facial expressions while students worked with an online tutoring system. Association rule mining identified that frustration was associated with activity in facial action units (AU) 1, 2 (inner and outer brow raiser) and 14 (dimpler); confusion was associated with AU 4 (brow lowerer), 7 (lid tightener) and 12 (lip corner puller). So, facial expressions are an interesting modality for detecting mental states.

**Postures** Regarding postures, Kapoor and Picard (2005) present research in which posture data was collected together with facial expressions and computer information while children solved an educational computer puzzle. Sensors in the chair were used to extract posture features (like leaning back, sitting upright) and activity level (low, medium, high). Posture information yielded the highest unimodal accuracy (82.52%) with an SVM for estimating interest (vs. uninterest). Performance was further improved by adding facial expression and

computer information in a multimodal Gaussian Process approach. We conclude that posture information and movement are an interesting source for estimating the users' mental state. We see potential for posture measurements in the office, as with the Kinect, recently an affordable 3D camera with skeleton detection has entered the market.

**Computer interactions** Finally, in some research, stress or emotions are estimated from computer interaction data. Vizer, Zhou, and Sears (2009) provide an overview of related work, and they present own work on the effect of stress on keystroke and linguistic features. Participants first performed a mentally or physically stressful task (e.g. remembering digits or exercising) and were then asked to write an email. They applied the following classifiers: decision tree, SVM, kNN, AdaBoost and artificial neural networks. They state that 'These techniques were selected primarily because they have been previously used to analyze keyboard behavior (e.g. kNN), or they have shown good performance across a variety of applications (e.g. SVM)' (p.878). To address individual differences, data was also normalized per participant by using baseline samples. Results indicate that stress can produce changes in typing patterns. With an accuracy of 75% kNN generated the best results for detecting cognitive stress (based on normalized features). In general, normalization improved the accuracy of all techniques, with an average increase of 13.1%. Vizer, Zhou, and Sears (2009) conclude: 'individual differences should be taken into consideration when developing techniques to detect stress, especially if cognitive stress is of interest' (p.879).

In the work by Khan, Brinkman, and Hierons (2008) the user's mood is inferred from their computer behavior. They aggregated computer activities within a 6 and 10 minute time window around mood ratings, and applied a correlation analysis. For 31% of the 26 participants they found significant correlations between keyboard/ mouse use and valence, and for 27% of the participants they found significant correlations with arousal. They further found that valence can better be predicted for users with more experience and less self-discipline, whereas arousal can better be predicted for users that are more dutiful.

Furthermore, Epp, Lippold, and Mandryk (2011) recognize 15 emotional states based upon keystroke dynamics. For classification they used decision trees as a 'simple and low-cost solution' (p. 719). They did not create user specific models due to the large variation in responses per user.

Finally, van Drunen *et al.* (2009) did research on computer interaction data as indicator of workload and attention. The participants performed a task in which they were asked to search for an item on a website. They found a correlation

between mouse data and heart rate variability. So, computer interactions may also be an interesting modality for detecting mental states.

**To conclude,** all four modalities have previously been used for mental state inference with some success, although most researchers collected data in a lab setting. Only some report on experiments that are more close to real-world situations. Several machine learning approaches are applied. In most cases, classification is used to distinguish 2 or more states. Sometimes, correlation analysis is performed to assess the strength of a relation. We aim to not only compare different classification approaches, but also apply regression models to make numerical predictions of e.g. mental effort. Several researchers find that individual differences play a role. In the models individual differences are not addressed or addressed by normalizing data per participant based upon a baseline sample. Making models for subgroups of similar users seems to be a new approach.

## 6.3 Dataset

To investigate which preferably physically unobtrusive and readily available sensors are most suitable to infer working conditions and mental states of knowledge workers, a data collection study was performed (Koldijk, Sappelli, Verberne, *et al.*, 2014)<sup>1</sup>. We created a realistic knowledge worker setting in which the effects of external stressors on subjective experience of task load, mental effort and emotions, as well as the effects on behavior could be investigated. 25 participants performed knowledge worker tasks, like report writing. To manipulate the experienced task load, we chose two stressors that are relevant in the knowledge worker context: interruptions by incoming emails and time pressure to finish a set of tasks before a deadline.

The following data was captured with sensors: computer interactions via a computer logging tool, facial expressions via a webcam, postures via a Kinect 3D camera and physiology via body sensors. The raw sensor data was preprocessed, yielding the SWELL-KW dataset (for more details see Koldijk, Sappelli, Verberne, *et al.* (2014)). This dataset contains preprocessed data, aggregated per minute, for all 25 participants. It contains the following features (see Table 6.1): 12 computer interaction features, 40 facial expression features, 88 body posture features and 3 physiology features. Per participant ca. 45 minutes of working under normal conditions were collected, ca. 45 minutes working with email interruptions and ca. 30 minutes working under time pressure. The feature

---

<sup>1</sup>Chapter 3

dataset is annotated with the conditions under which the data was collected. The possibly chaotic minutes at the very beginning and very end of each condition were removed. The dataset contains 149 features and 2688 instances in total (on average 107 minutes data per participant).

Moreover, the dataset includes a ground truth for subjective experience (see Table 6.2). This was assessed by means of validated questionnaires on task load (NASA-TLX, Hart and Staveland (1988)), mental effort (RSME, Zijlstra and van Doorn (1985)), emotion (SAM, Bradley and Lang (1994)) and perceived stress (own visual analog scale). 25 participants each rated 3 working conditions, which yielded 75 ground truth ratings in total. Note that one rating corresponds to 30-45 minutes of working behavior data. In our dataset therefore, we repeatedly annotated each row of one minute data with the ground truth of that condition. In previous work (Koldijk, Sappelli, Neerincx, *et al.*, 2013) we found that the stressor working conditions indeed affected subjective experience, see Figure 6.2. Therefore, we can use this dataset for user state modeling in stress related terms.

In our work we make the following assumptions: 1) Facial expressions, postures and physiology were reliably inferred from the raw sensor data. 2) Aggregated data over 1 minute yields valuable information. 3) Subjective ratings provide a good ground truth. 4) The subjective rating given to the entire condition can be used as ground truth for each separate minute.

Table 6.1: SWELL-KW feature dataset. Data is preprocessed and aggregated per minute. The dataset contains 149 features and 2688 instances.

Modality (#features)	Feature type (#features)
Computer interactions (18)	Mouse (7)
	Keyboard (9)
	Applications (2)
Facial expressions (40)	Head orientation (3)
	Facial movements (10)
	Action Units (19)
	Emotion (8)
Body postures (88)	Distance (1)
	Joint angles (10)
	Bone orientations (3x11)
	(as well as stdv of the above for amount of movement (44))
Physiology (3)	Heart rate (variability) (2)
	Skin conductance (1)

Table 6.2: Subjective experience data (3 ratings per participants). Average values for the Neutral, Interruption and Time pressure condition can be found in the last 3 columns.

Type	Feature	N	I	T
TaskLoad	MentalDemand (0: low - 10: high)	4.9	5.4	4.9
	PhysicalDemand	1.9	2.3	2.7
	TemporalDemand	5.7	5.9	7.1
	Effort	5.2	5.9	6.1
	Performance	4.8	6.1	6.0
	Frustration	3.5	3.6	3.5
Mental Effort	MentalEffort (0: no - 10: extreme effort)	5.5	6.5	6.3
Emotion	Valence (1: unhappy - 9: happy)	4.8	5.7	5.3
	Arousal (1: calm - 9: excited)	3.3	3.9	4.6
	Dominance (1: submissive - 9: dominant)	5.2	6.2	5.9
Stress	Perceived stress (0: not - 10: very stressed)	2.9	3.2	3.8

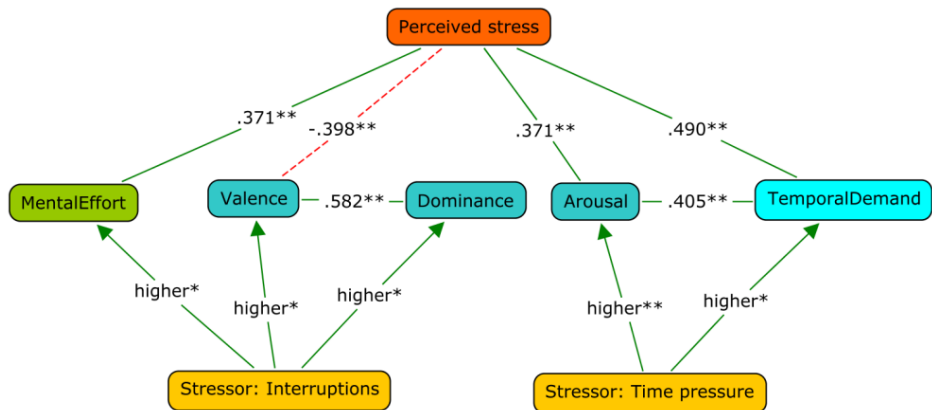


Figure 6.2: Bottom part: The stressors affected several aspects of subjective experience. Top part: These aspects of subjective experience correlated with perceived stress. (\* significant at the .05-level, \*\* significant at the .001-level.)

## 6.4 Using several unobtrusive sensors to detect stress in office environments

In this section, we aim to answer our first main research question: Can we distinguish stressful from non-stressful working conditions, and can we estimate mental states of office workers by using several unobtrusive sensors? To distin-

guish stressful working conditions from neutral ones, we use *classification models* (Section 6.4.1). Moreover, to estimate mental states, like the amount of mental effort and stress, we use *regression models* (Section 6.4.2). We compare the performance of several machine learning approaches and also investigate which modalities (computer interactions, facial expressions, posture, physiology) and which particular features are most suitable.

### 6.4.1 Inferring the working condition

We first investigate whether we can distinguish stressful from non-stressful working conditions, i.e. we try to predict whether a minute of data is from the normal working condition (N, 1028 instances), or from a stressful working condition (T&I, i.e. time pressure (664 instances) or email interruptions (996 instances)).

#### Comparison of different classifiers

We start with our first subquestions: Which modeling approaches are most successful? We selected the following types of classifiers (we used their implementations in the machine learning toolkit Weka (Hall *et al.*, 2009)):

- Nearest neighbors: IBk (uses euclidean distance, we tested the following number of neighbors: 1, 5, 10, 20), K-star (uses entropic distance)
- Bayesian approaches: Naive Bayes, Bayes Net
- Support vector machines (SVM): LibSVM (we tested the following kernels: linear, polynomial, sigmoid, radial basis function)
- Classification trees: J48 (decision tree), Random Forest
- Artificial neural networks: Multilayer perceptron

(For a comprehensive and concise explanation of the different approaches, we refer to Novak, Mihelj, and Munih (2012)). Based on the presented related work we expect that nearest neighbors, trees and SVM perform well.

As features we always used the entire SWELL-KW dataset with features on 4 modalities: computer interactions, facial expressions, physiology and postures. For nearest neighbor, SVM and neural network we first standardized the features to 0 mean 1 std (as was done in Caruana and Niculescu-Mizil, 2006, for Bayes and trees scaling of the features is not necessary). We evaluated our models on accuracy, i.e. the percentage of correctly classified instances. We used 10-fold cross-validation. (This means that the data is randomly split into 10 equally

large folds. 90% of the data (9 folds) is then used for training the model, the remaining fold is used for testing. This is repeated 10 times, and scores are averaged.)

The results for the different classifiers are presented in Table 6.3. In general, a performance of about 90% accuracy can be reached in distinguishing neutral from stressful working conditions based on sensor data, which is high.

Regarding the different classification approaches, we find the following: The *Bayesian approaches* score only somewhat above baseline (61.7600%): naive Bayes (64.7693%), and Bayes net (69.0848%). This means the data is not well-modeled in terms of chance distributions, which is what we expected. Regarding the *nearest neighbor classifiers*, KStar does not work well (65.8110%). However, IBk (which uses an euclidean distance measure) reaches a really good performance with 10 neighbors (84.5238%). Looking up nearby data points thus seems to work for this dataset, which was expected. Also as expected, *classification trees* seem to work on our data: decision tree (78.1994%), and random forest (87.0908%). The advantage of a decision tree approach is that the model is insightful. The *artificial neural network* also yields a good result: 88.5417%. However, it takes very long to train a neural network model (1 hour in our case). Finally, the best results were obtained with an SVM (using a radial basis function kernel): 90.0298%. That SVM perform well was also found in previous work.

Classifier	Accuracy
BASELINE Majority class: Stressful (I&T)	61.76%
Naive Bayes	64.7693%
K-star (nearest neighbor with entropic distance)	65.8110%
Bayes Net	69.0848%
J48 (decision tree)	78.1994%
IBk (nearest neighbor with euclidean distance), 10 neighbors	84.5238%
Random Forest	87.0908%
Multilayer perceptron (neural network)	88.5417%
LibSVM (support vector machine), radial basis function kernel	90.0298%

Table 6.3: Comparison of classifiers. Predict working conditions (N vs. I&T) from 4 modalities (Computer, Facial, Physiology, Posture).

### Comparison of different feature sets

Now, we address our second subquestion: Which modalities/ features provide the most useful information? We continued our analyses with the SVM classifier, as this performed best.



First, we tested the following subsets: 1 modality (computer, facial, physiology or posture), a combination of 2 modalities, 3 modalities or all 4 modalities. We hypothesize that combining different modalities improves classification performance. However, for an office setting, excluding sensors to yield a minimal set-up would be preferable.

Our results are presented in Table 6.4. When using only a single modality, posture features yield best performance (83.4077%). When adding a second modality, facial features yield improvements (88.6905%). Only minor improvements can be reached when adding a third modality (physiology: 89.2857% or computer features: 89.1369%), and only a slightly higher accuracy is reached when using all four modalities (90.0298%). So, as expected, combining more than one modality improved performance, although the gains due to additional modalities are modest. In an office setting thus the most important modality to use would be posture, possibly combined with facial information.

We also ran feature selection in Weka<sup>2</sup>, yielding a subset of the 26 best features for predicting the working condition. This set includes 17 posture features (13 features related to sitting posture, 4 related to body movement), 5 facial expression features (LidTightener, rightEyebrowRaised, Dimpler, looking surprised, and happy), 2 physiological features (heart rate, and skin conductance) and 2 computer interaction features (applicationChanges, and leftClicks). This confirms our hypothesis that features from different modalities can complement each other. With only these 26 features, still a performance of 84.5238% can be reached.

## Conclusions

In this section we investigated the research question: Can we distinguish stressful from non-stressful working conditions by using several unobtrusive sensors? We found that a performance of about 90% accuracy can be reached, which is reasonably high. SVM, neural networks and random forest approaches yield the best performance. Also the rather simple nearest neighbor and decision tree approaches seem to provide reasonable accuracy. On the other side, Bayesian approaches seem less suitable for this data. With respect to the most useful modalities we find that posture yields most valuable information to distinguish stressor from non-stressor working conditions. Adding information on facial expressions can further improve performance. Computer interactions and physiology, however, showed no gains, which was unexpected given the presented related work on stress recognition based upon these modalities.

---

<sup>2</sup>CfsSubsetEval with BestFirst search

Features	Accuracy
BASELINE: ZeroR (majority class I&T)	61.76%
1 modality (Computer, 18 features)	65.5134%
1 modality (Physiology, 3 features)	64.0997%
1 modality (Facial, 40 features)	75.4092%
<b>1 modality (Posture, 88 features)</b>	<b>83.4077%</b>
2 modalities (Computer & Physiology, 21 features)	67.8943%
2 modalities (Computer & Facial, 58 features)	79.1295%
2 modalities (Facial & Physiology, 43 features)	79.9479%
2 modalities (Posture & Computer, 106 features)	83.7798%
2 modalities (Posture & Physiology, 91 features)	83.7798%
<b>2 modalities (Posture &amp; Facial, 128 features)</b>	<b>88.6905%</b>
3 modalities (Computer & Facial & Physiology, 61 features)	81.2872%
3 modalities (Posture & Computer & Physiology, 109 features)	84.0402%
3 modalities (Posture & Facial & Computer, 146 features)	<b>89.1369%</b>
3 modalities (Posture & Facial & Physiology, 131 features)	<b>89.2857%</b>
4 modalities (Computer & Facial & Physiology & Posture, 149)	<b>90.0298%</b>
Only 26 best features	84.5238%

Table 6.4: SVM with radial basis function kernel. Comparison of using feature subsets to predict working conditions (N vs. I&T).

### 6.4.2 Predicting mental states

In the data collection experiment also information of subjective experience was collected after each working condition. We have information on: perceived stress, mental effort, emotion (i.e. valence, arousal, dominance), and task load (i.e. mental demand, physical demand, temporal demand, performance, effort, frustration). We address the following question now: Can we estimate mental states of office workers by using several unobtrusive sensors?

#### Predicting different subjective variables

First of all, we investigated which subjective variable can best be predicted from our sensor data. Our main aim is to infer stress from sensor data. However, other relevant subjective variables may be more directly related to the recorded sensor data.

We used Weka to train linear regression models. For evaluation, we used the correlation between the predicted values and the true values. We also used RMSE (root mean squared error) as measure of error in the predictions. We applied 10fold cross-validation again. As features we always used the entire SWELL-KW dataset with features on 4 modalities: computer interactions, facial

expressions, physiology and postures.

The results for predicting different subjective variables from sensor data are presented in Table 6.5. The variable mental effort can best be predicted from our sensor data, yielding a reasonably high correlation of 0.7920 and the lowest RMSE. Other related subjective variables (perceived stress, arousal, frustration, valence, task load and temporal demand) can all be predicted equally well, with a lower correlation of around .7 (and a worse RMSE). This shows that mental effort is easier to read from facial expressions, posture, computer interactions and physiology, than e.g. stress.

Prediction variable	Correlation	RMSE
Mental effort	0.7920	0.6115
Valence	0.7139	0.7024
Arousal	0.7118	0.7044
Frustration	0.7117	0.7048
Perceived stress	0.7105	0.7054
Task load	0.6923	0.7241
Temporal demand	0.6552	0.7592

Table 6.5: Linear regression. Predicting different subjective variables from our 4 modalities (computer, facial, physiology and posture features). Data was standardized to 0 mean 1 std, for fair comparison of RMSE.

**Best features for each of the subjective variables** We also ran feature selection in Weka, to see which set of features is used to predict a specific subjective variable (see Tables 6.6 and 6.7). In general, most often several facial and posture features are selected for predicting mental states, sometimes combined with a physiological feature. It is interesting to see that the algorithm selects different specific features for different subjective variables. These selected features seem to make sense: e.g. skin conductance to predict stress and frustration, or the amount of error keys to predict arousal.

### Comparison of different regression models

As the variable mental effort seemed to be best predictable from our sensor data we focus on this subjective variable for the remainder of our analyses. We now address our first subquestion again: Which modeling approaches are most successful? We selected the following types of regression models (we used their implementations in the machine learning toolkit Weka):

- Linear regression

Type	MentalEffort	Stress
Facial	sad surprised scared rightEyebrowLowered gazeDirectionLeft Au06_CheekRaiser Au07_LidTightener Au10_UpperLipRaiser Au17_ChinRaiser	surprised Au05_UpperLidRaiser Au06_CheekRaiser Au15_LipCornerDepressor Au26_JawDrop Au43_EyesClosed
Posture	2 posture features (leaning, left shoulder)  1 movement feature (right elbow)	3 posture features (left shoulder, left wrist, head) 2 movement features (average body movement, right upper arm)
Computer	-	-
Physiology	-	skin conductance level

Table 6.6: Feature selection for the subjective variables stress and mental effort. (CfsSubsetEval with BestFirst, features were standardized to 0 mean and 1 stdv).

- Nearest neighbors: IBk (uses euclidean distance, we tested the following number of neighbors: 1, 5, 10), K-star (uses entropic distance)
- Support vector machine: SMOreg (we tested the following kernels: polynomial, radial basis function)
- Regression trees: REPTree (regression tree, i.e. sample mean at each leaf), M5P (model tree, i.e. function at each leaf)
- Artificial neural networks: Multilayer perceptron

We expected that a simple linear regression or tree approach, or the more complex SVM would work well with our data. As features we always used the entire SWELL-KW dataset with features on 4 modalities: computer interactions, facial expressions, physiology and postures.

The results on predicting mental effort from sensor data are presented in Table 6.8. Several models reach reasonable performance far better than baseline. The simple linear regression model reaches a comparable performance to the more complex support vector machine (correlation of 0.7920 vs. 0.7990; RMSE of 0.6115 vs. 0.6035). Furthermore, a model tree approach seems to work best

Type	Arousal	Frustration
Facial	angry scared zHeadOrientation rightEyeClosed leftEyebrowRaised Au06_CheekRaiser Au09_NoseWrinkler Au10_UpperLipRaiser Au25_LipsPart	(quality) gazeDirectionRight Au05_UpperLidRaiser Au09_NoseWrinkler Au14_Dimpler Au15_LipCornerDepressor Au23_LipTightener Au43_EyesClosed
Posture	6 posture features (head, shoulder center, left shoulder, left upper arm, left wrist, right upper arm) 1 movement feature (right lower arm)	4 posture features (left shoulder, left upper arm, left wrist, right upper arm)  5 movement features (leaning, left upper arm, left wrist, right upper arm, right lower arm)
Computer	nErrorKeys	nRightClicked
Physiology	heart rate variability	skin conductance level

Table 6.7: Feature selection for the subjective variables arousal and frustration. (CfsSubsetEval with BestFirst, features were standardized to 0 mean and 1 stdv).

with a correlation between predicted and real mental effort of 0.8221 and the lowest RMSE (0.5739). This is in line with our expectations.

### Comparison of different feature sets

For its good performance and speed, we decided to continue our analyses with the model tree. We now address our second subquestion again: Which modalities/ features provide the most useful information?

We tested the following subsets: 1 modality (computer, facial, physiology or posture), a combination of 2 modalities, 3 modalities or all 4 modalities. We hypothesize that combining different modalities improves classification performance. However, for an office setting, excluding sensors to yield a minimal set-up would be preferable.

Our results are presented in Table 6.9. When using only a single modality, facial features yield the best performance with a correlation between predicted and true values of 0.8091. When adding a second modality, only posture features yield a slight improvement (0.8300). No real improvement is gained when

Classifier	Correlation	RMSE
ZeroR (baseline)	-0.0703	1.0004
Kstar (nearest neighbor with entropic distance)	0.5875	0.9104
IBk (nearest neighbor with euclidean distance), 5 neighbors	0.7330	0.7229
REPTree (regression tree)	0.7577	0.6534
Multilayer Perceptron (neural network)	0.7763	0.7064
Linear regression	0.7920	0.6115
SMOreg (SVM), with radial basis function kernel	0.7990	0.6035
M5P (model tree)	0.8221	0.5739

Table 6.8: Comparison of regression models. Predict mental effort from 4 modalities (Computer, Facial, Physiology, Posture). Data was standardized to 0 mean 1 std, for fair comparison of models.)

Features	Correlation
BASELINE: ZeroR	-0.0637
1 modality (Computer, 18 features)	0.1545
1 modality (Physiology, 3 features)	0.5715
<b>1 modality (Facial, 40 features)</b>	<b>0.8091</b>
1 modality (Posture, 88 features)	0.5896
2 modalities (Computer & Physiology, 21 features)	0.5527
2 modalities (Computer & Facial, 58 features)	0.8027
2 modalities (Facial & Physiology, 43 features)	0.7891
2 modalities (Posture & Computer, 106 features)	0.6254
2 modalities (Posture & Physiology, 91 features)	0.7644
<b>2 modalities (Posture &amp; Facial, 128 features)</b>	<b>0.8300</b>
3 modalities (Computer & Facial & Physiology, 61 features)	0.7909
3 modalities (Posture & Computer & Physiology, 109 features)	0.7718
3 modalities (Posture & Facial & Computer, 146 features)	0.8182
3 modalities (Posture, Facial, Physiology, 131 features)	0.8295
4 modalities (Computer, Facial, Physiology, Posture, 149)	<b>0.8309*</b>
<b>Only 25 best features</b>	<b>0.8416</b>

Table 6.9: Decision tree (M5P). Comparison of feature subsets to predict mental effort. (\* slightly differs from result in Table 6.8, as for the tree here non-standardized features were used.)

adding more modalities. Contrary to our expectations, it seems best to merely use facial features, or just add posture information to predict mental effort.

We also ran feature selection in Weka<sup>3</sup>, yielding a subset of the 25 best features for predicting mental effort with a model tree. This subset includes 10 facial features (BrowLowerer, UpperLidRaiser, LidTightener, UpperLipRaiser, looking sad, angry, left and right eyebrows lowered, xHeadOrientation, and gazeDirectionLeft), 9 posture features (5 features related to sitting posture, and 4 related to body movement), 2 physiological features (heart rate, and skin conductance), and 4 computer interaction features (right click, double click, direction keys, and error keys). This confirms our hypothesis that features from different modalities can complement each other. With only these 25 features a performance of 0.8416 can be reached, which is slightly better than the best accuracy, which was reached with all 149 features (0.8309). Although for a real-world office setting it might be more interesting to restrict the setup to only facial expressions, which worked well as single modality (0.8091).

## Conclusions

In this section we investigated the research question: Can we estimate mental states of office workers by using several unobtrusive sensors? First of all, we found that mental effort seems to be the variable that can be best predicted from our sensor data (better than stress, arousal, frustration, valence, task load or temporal demand). A comparison of different regression models showed that a performance of 0.8221 can be reached with a model tree, which is reasonably high. Also linear regression models, or SVMs provide good accuracy. With respect to the most useful modalities we find that facial expressions yield most valuable information to predict mental effort. Adding information on posture can slightly improve performance.

## 6.5 Taking into account individual differences

Until now, we built one generic model over all users. Now, we address our second main research question: How important are individual differences? First we investigate the role of individual differences regarding working conditions (Section 6.5.1), then regarding mental states (Section 6.5.2). Finally, we investigate the performance of models that were built for specific user groups (Section 6.5.3).

---

<sup>3</sup>Wrapper for M5P with BestFirst search

### 6.5.1 Individual differences regarding working condition

We start with investigating the role of individual differences in the classification problem, i.e. predicting whether a data point is from the normal working condition, or from a stressful working condition.

**Participant ID as feature** To investigate in how far the models benefit from participant information, we add the participant ID as feature to the dataset. Recall that a SVM predicting stressor vs. non-stressor working conditions based on all 4 modalities reached a performance of 90.0298%. When we add participant ID to the set of features, the SVM reaches a comparable performance: 89.6577%. This means that knowledge on which specific user needs to be classified, yields no valuable information to the model.

We performed feature selection to test the relative importance of the participant ID feature. We find that the participant ID feature has a very low information gain and gain ratio, and is not selected in the subset of best features for predicting the working condition.

As decision trees provide most insight, we decided to also apply this model to our data. When we inspect the built decision tree, we see that participant ID is a feature which occurs relatively late in the tree. It is thus not the case that the model builds different sub-trees for different users. However, it is the case that towards the end of a branch describing particular behavior, a split-point based on the participant can be found. So the same behavior may need to be interpreted differently, depending on the user at hand. This may indicate that different users display stress in different ways.

**Test on unseen user** Furthermore, we investigated how general models perform on new, unseen users. Therefore, we trained a SVM on 24 user's data and test it on a left out, unseen, user (leave-one-subject-out cross validation). What we see is a drop in performance. It differs per user how bad this drop is. Remember that 10-fold cross-validation yielded a performance of 90.0298%. When testing on an unseen user, the model reaches an average accuracy of only 58.8887%. (Recall that the baseline for our dataset was 61.76%). The worst performance was reached for participant 21, namely 37.5000%. The best performance was reached for participant 2, namely 88.3495%. The standard deviation between performances on different users was with 11.6376%-points relatively high. Whether a model performs well on a new, unseen user may depend on the similarity of the new user to previous users.



**Conclusions** With respect to distinguishing stressor from non-stressor working conditions, we see that information on the particular participant does not improve classification accuracy. We also find, that the participant ID is not important enough to be selected as one of the best features for this classification task. In the decision tree, the participant ID only appears late in the branches, helping to interpret the same behavior differently for different users. When we test a generic model on an unseen user, we see a drop in performance. It differs per user how big this drop is. This may depend upon the similarity of the new user to previous users.

### 6.5.2 Individual differences regarding mental states

We now investigate the role of individual differences in the regression problem, i.e. predicting the amount of mental effort based on sensor data.

**Participant ID as feature** To investigate in how far the models benefit from participant information, we add the participant ID as feature to the dataset again. Recall that a decision tree predicting mental effort based on all 4 modalities reached a performance of 0.8221 (RMSE was 0.5739). When we add participant ID to the set of features, the decision tree reaches a higher performance: 0.9410 (RMSE is 0.3383). This means that knowledge on the specific user yields valuable information to the model.

We performed feature selection to test the relative importance of the participant ID feature. We find that the participant ID is selected in the subset of 13 best features for predicting mental effort (besides 9 facial expression and 3 posture features).

When we inspect the built decision tree we see that the participant ID is included in the regression formulas: for groups of users specific weights are added or subtracted.

**Test on unseen user** Furthermore, we also investigated how generic models perform on new, unseen users. Therefore, we trained a decision tree model on 24 user's data and test it on a left out, unseen, user (leave-one-subject-out cross validation). What we see is a drop in performance. Remember that 10-fold cross-validation yielded a performance of 0.8221 (RMSE 0.5739). When testing on an unseen user, the model reaches an average correlation of only 0.0343 (average RMSE: 1.1684). (Recall that the baseline for our dataset was a correlation of -0.0703, with an RMSE of 1.0004). Predicting the mental effort of an unseen user is thus difficult. In terms of correlation, the worst performance was reached for participant 20, namely a negative correlation of -0.4311. The best performance

was reached for participant 5, namely a correlation of 0.7241. The standard deviation between performances on different users was with 0.2800 reasonably high. Whether a model performs well on a new, unseen user may depend on the similarity of the new user to previous users.

**Conclusions** With respect to estimating mental states, we see that information on the particular participant improves the mental effort estimates. The participant ID is even important enough to be selected as one of the best features for this regression task. In the regression formulas we see that specific weights are added or subtracted for groups of users. A general model tested on a new user does not perform well. This suggests that especially for the task of estimating mental states it makes sense to address individual differences.

### 6.5.3 Addressing individual differences

Finally, we investigate how individual difference can be addressed. We test whether the performance of the regression models for a single modality can be improved when distinct models are made for groups of similar users.

**Clustering of users** In previous work (Koldijk, Bernard, *et al.*, 2015) we clustered users into groups, with respect to their average level of: computer activity, facial expressions or postures. Hierarchical clustering was used to reveal the amount of clusters ( $k$ ) in the data and then k-means clustering was applied. We addressed each sensor separately and found that for each sensor the users were grouped differently. This yielded, for each modality, particular groups of similar users.

**Computer activity groups** We found that, based on average computer activity, 2 groups of users can be discriminated: the ‘writers’ (16 participants (PP), many keystrokes) and the ‘copy-pasters’ (9 PP, much mouse activity and special keys). Recall that the performance of a decision tree with only computer activity features for predicting mental effort yielded a performance of 0.1545. When training and validating a model only on the ‘writers’, we find an equal correlation of 0.1668 for predicting mental effort. When training and validating a model only on the ‘copy-pasters’, we find a higher correlation of 0.3441.

Furthermore, we applied feature selection to find the features most predictive of mental effort for both groups. For ‘writers’, the best features to predict mental effort are: amount of right clicks and scrolling. For ‘copy-pasters’, however, the best features to predict mental effort are: the amount of dragging, shortcut keys, application and tabfocus changes, as well as the error key ratio.

**Facial action unit groups** We found that, based on average facial action unit activity, 3 groups of users can be discriminated: The ‘not very expressive’ ones (16 PP), the ‘eyes wide & mouth tight’ group (3 PP), and the ‘tight eyes & loose mouth’ group (6 PP). Recall that the performance of a decision tree with only facial features for predicting mental effort yielded a performance of 0.8091. When training and validating a model only on the ‘not very expressive’, we find a slightly worse correlation of 0.7892. When training and validating a model only on the ‘eyes wide & mouth tight’ group, we find an equal correlation of 0.8091. When training and validating a model only on the ‘tight eyes & loose mouth’ group, we find a higher correlation of 0.8742.

Furthermore, we applied feature selection to find the features most predictive of mental effort for both groups. For ‘not very expressive’ users, the best features to predict mental effort include the action units: Dimpler and LipsPart. For ‘eyes wide & mouth tight’ users, the best features to predict mental effort include the action units: LidTightener, UpperLipRaiser, LipCornerPuller and ChinRaiser. For ‘tight eyes & loose mouth’ users, the best features to predict mental effort include the same action units, but additionally also: BrowLowerer, Dimpler, MouthStretch and EyesClosed.

**Body movement groups** We found that, based on average body movement, 3 groups of users can be discriminated: the group that ‘sits still & moves right arm’ (5 PP), the group that ‘moves body a lot & wrist less’ (6 PP) and the group that ‘moves average’ (14 PP). Recall that the performance of a decision tree with only posture features for predicting mental effort yielded a performance of 0.5896. When training and validating a model only on the group that ‘sits still & moves right arm’, we find a higher correlation of 0.7564. When training and validating a model only on the group that ‘moves body a lot & wrist less’, we find a higher correlation of 0.8488. When training and validating a model only on the group that ‘moves average’, we find a higher correlation of 0.6917.

**Conclusions** When we train models on particular subgroups of similar users, (in almost all cases) a specialized model performs equally well or better than a general model. With respect to computer activity, the model for ‘writers’ performs similar to a general model, whereas a model for ‘copy-pasters’ outperforms our general model. With respect to facial activity, the model for ‘not very expressive’ users performs slightly worse than a general model. However, the model for the ‘eyes wide & mouth tight’ group performs the same as our general model. And the model for the ‘tight eyes & loose mouth’ group really outperforms our general model. Finally, with respect to posture, all models for the sub-groups really outperform our general model. We also find that for differ-

ent user groups, different features are selected. To apply models for subgroups of users in office settings, data of an initialization phase may be necessary to categorize a user into one of the subgroups based upon his average behavior.

## 6.6 Conclusions and Discussion

In this chapter we investigated different machine learning approaches to infer working conditions and mental states from a multimodal set of sensor data (computer logging, facial expressions, posture and physiology).

We addressed two methodological and applied machine learning challenges: 1) Detecting work stress using several (physically) unobtrusive sensors. We first answered the following research question: Can we distinguish stressful from non-stressful *working conditions* by using several unobtrusive sensors? We found that a performance of about 90% accuracy can be reached. SVM, neural networks and random forest approaches work best. Also the rather simple nearest neighbor and decision tree approaches seem to provide reasonable accuracy. With respect to the most useful modalities we find that posture yields most valuable information to distinguish stressor from non-stressor working conditions. Adding information on facial expressions further improves performance.

Moreover, we answered the research question: Can we estimate *mental states* of office workers by using several unobtrusive sensors? Mental effort seems to be the variable that can be best predicted from our sensor data (better than e.g. stress). A comparison of different regression models showed that a performance of 0.8221 can be reached. Model trees yield the best performance. Also linear regression models, or SVMs provide good accuracy. With respect to the most useful modalities we find that facial expressions yield most valuable information to predict mental effort. Adding information on posture can slightly improve performance.

Then, we addressed the second methodological and applied machine learning challenge: 2) taking into account individual differences. We first answered the research question: How important are individual differences? With respect to distinguishing stressor from non-stressor working conditions, we see that information on the participant is not important enough to be selected as one of the best features. In the decision tree, the participant ID only appears late in the branches. When we test a generic model on an unseen user, we see a drop in performance. It differs per user how big this drop is. This may depend upon the similarity of the new user to previous users. With respect to estimating mental states, we see that information on the participant is important enough to be selected as one of the best features. In the regression formulas we see that specific weights are added or subtracted for groups of users. We further find

that a general model tested on a new user does not perform well. This suggests that especially for the task of estimating mental states it makes sense to address individual differences. It should be investigated in future work why individual differences seem to play a bigger role in estimating mental states than in distinguishing neutral from stressor working conditions.

Finally, we answered the research question: Can we improve performance by building personalized models for particular user groups? When we train models on particular subgroups of similar users, (in almost all cases) a specialized model performs equally well or better than a general model. Especially with respect to facial activity, the model for the group 'tight eyes & loose mouth' really outperforms our general model. Also, with respect to posture, all models for the sub-groups really outperform our general model. We also find that for different user groups, different features are selected.

We have to note that, a good approach to address individual differences could also be to build models for single users. However, the amount of data we had available per participant here was not enough (only 3 different subjective ratings).

Our work was based on several assumptions, on which we will comment now: 1) Facial expressions, postures and physiology were reliably inferred from the raw sensor data. The data that we used here, was captured in a realistic office setting in an experimental context, which means that the quality of all recordings was high. In a real-world office setting recordings may be more noisy, e.g. facial expression recognition may be less reliable with bad lighting, or when the user is not positioned well in front of the camera. Moreover, specialist equipment for capturing physiology was used here. In real-world settings, devices like smart measuring watches may provide less reliable data.

2) Aggregated data over 1 minute yields valuable information. There are many potential choices on how to handle the aspect of time. Here we chose to aggregate data per minute and classify each minute separately. Alternatively, different time-frames can be considered. Moreover, a model that takes into account relations between time-frames may be suitable. Finally, we have to note that consecutive minutes may be very similar. A random split for cross-validation may thus contain test cases that are very similar to training cases. This effect may be stronger when data of less different participants is used. On the one side, learning from similar examples is exactly what machine learning aims to do. On the other side, we have to note that such very similar minutes may cause high accuracy in evaluation.

3) Subjective ratings provide a good ground truth. There is debate on whether subjective ratings provide a good ground truth. An alternative would be to use e.g. physiology as ground truth for stress. Here we chose to use sub-

jective ratings, because we expected that physiological stress reactions in office setting would not to be very strong.

3) The subjective rating given to the entire condition can be used as ground truth for each separate minute. It may be argued that stress experienced due to time pressure or incoming emails, may become stronger as the deadline comes closer or more emails have interrupted the user. Therefore, one could argue to only use the data from the last part of the condition. As we do not have too much data per participant, however, we decided to include the entire condition. Moreover, behavior may fluctuate over time. Not each minute may include signs of stress, whereas others do. The good accuracy of the classification and regression approaches, however, indicates that not too much noise was introduced into our models in this way.

To conclude, the four modalities were successfully used in an office context. Several classification and regression models were compared to find the most suitable approach. We also investigated which modalities and features were most informative. Besides applying generic models, we investigated the role of individual differences. We showed how models for subgroups of similar users can be made.

We have to note that we did all analyses on one specific dataset, the SWELL-KW dataset. Our results may be dependent on specific characteristics of this dataset. First of all, the participants' behavior is dependent on the specific tasks we gave them. This may be especially reflected in our computer interaction data: part of the interactions we record are related to the tasks of writing reports and making presentations.<sup>4</sup>. Note however, that the tasks themselves stayed the same for all 3 working conditions, the only thing that may have changed due to our stressors is the manner of working. Computer logging can capture, besides task related aspects, general computer interaction characteristics that change under stress, e.g. a faster typing speed or quicker window switching. These may generalize to other office working contexts, and are thus independent of our chosen tasks. Second, the specific stressors we chose, time pressure and email interruptions, may have a specific influence on the participants behavior, like a quicker work pace or more effort to concentrate on the task at hand. This may explain why stress itself was harder to predict from our sensor data. Mental effort may be more closely related to the behavior displayed under these stressors. Finally, the (behavioral) signs of stress may be intertwined with a specific way of working. An interesting question is in how far the results found in our knowledge work context hold for stress detection in general. Throughout our

---

<sup>4</sup>We also did research on automatic task recognition to investigate the manner of working. A visualization of these results can be seen here: [http://cs.ru.nl/~skoldijk/Visualization/ExperimentBrowser/Generic/Gantt\\_and\\_Numeric2.html](http://cs.ru.nl/~skoldijk/Visualization/ExperimentBrowser/Generic/Gantt_and_Numeric2.html)

analyses, the facial expression features proved to be well suited. We think that facial expressions are a rather general expression of mental effort, which holds among different contexts. Moreover, posture features proved suitable. These have a little less generalizability, as the participant's posture is clearly related to the task of working behind a computer. However, it may be independent of the exact tasks that are performed. All in all, we can conclude that in future work our analyses should be applied to another dataset to prove the generalizability of the findings presented in this paper.

In general, the affective computing community often uses (black-box) machine learning algorithms to classify sensor data into mental states. In this work, we also investigated which behavior (e.g. typical facial activity, leaning forward, sitting still) that can be captured with sensors, is indicative of mental states related to stress. We see potential in building inference models that use an intermediate behavioral layer. This is in line with what S. Scherer *et al.* (2012) propose. We expect that a model with a more abstract intermediate behavior layer is more robust to individual differences and generalizes better over different users. This should be investigated in future work. In previous work (Koldijk, Bernard, *et al.*, 2015), we e.g. applied a supervised Self-organizing Map (SOM) to find typical facial expressions related to high mental effort, which could be used as intermediate behavior layer. The same analyses could be applied to posture or computer interaction data, to yield more behavioral patterns for a layered model.

# Human-centered development method for effective and privacy-friendly context aware support systems - A case study

*This chapter presents an improved human-centered development method, tailored to the design of effective and privacy-friendly context aware support systems. We refined the 'situated cognitive engineering' methodology on two aspects: 1) defining the to be sensed 'context' during the requirements engineering process, and 2) addressing functional and non-functional requirements coherently. We particularly focus on analyzing user concerns, complementing the analysis with a privacy impact assessment, and suggesting ways to address privacy in CAS. We present a case study: the development of the context aware SWELL system, which collects various contextual information to provide support for well-being at work. We found that users' data collection preferences are diverse and are based on individual trade-offs regarding privacy. The described human-centered development method supports balanced design decisions and requirements specifications on data collection. The resulting CAS is grounded in theory, takes into account technical possibilities and specifically addresses user concerns regarding data collection and privacy. We therefore recommend this approach for the development of other CAS.*

---

This chapter is based on Koldijk, Siljee, Neerinx & Kraaij. "Human-Centered Development Method for Effective and Privacy-Friendly Context Aware Support Systems - A Case Study". Submitted to: Cognition, Technology & Work.



## 7.1 Introduction

Advances in sensing, the rise of smartphones and mobile internet, together with trends such as ubiquitous user modeling and personalization have led to new possibilities in the area of supporting context-aware systems (CAS) and such applications are now very common. Context-awareness for applications has been broadly defined as the use of environmental elements by applications to personalize their service for the user (Abowd *et al.*, 1999). A simple example is given by a navigation application: you tell the application where you want to go, the explicit data, while the application obtains your current location from the mobile device, which is then the contextual data. Until now, many CAS make use of the users' external context, like location, temperature, sound, time or surrounding users (Hong, Suh, and Kim, 2009).

From a technological perspective, we can build even richer user models, including the users' internal context, like mental states or emotions. Porayska-Pomsta *et al.* (2013) showed that several mental states (stressed, embarrassed, ill-at-ease, bored, focused, hesitant, relieved) can be recognized during job interviews based on social cues sensed by a Kinect 3D camera. Researchers were able to infer personality traits from email messages (Shen, Brdiczka, and Liu, 2013). Moreover, the field of affective computing aims to infer emotions from speech, facial expressions, body movements (Tao and Tan, 2005) or physiological signals like heart rates and skin conductance (Nasoz *et al.*, 2003), which works to a reasonable extend under controlled settings. With new technology, like smart phones and smart watches entering the market, sensor data becomes more and more easy to collect from users. As a result, new possibilities for services arise that take internal context into account, e.g. offering adaptive support to learners (Salmeron-Majadas, Santos, and Boticario, 2013) or providing a personalized music player that responds to affect (Janssen, van den Broek, and Westerink, 2012). In the area of CAS, there has been considerable work in the domains of smart spaces (homes, hospitals, class rooms), tour guides, information systems, communication systems, m-commerce and web services (Hong, Suh, and Kim, 2009).

In this work, our case study is the development of the SWELL<sup>1</sup> system, a CAS for stress reduction, which is a new domain with interesting challenges around data collection:

The SWELL system provides information about working behavior to help employees self-manage their well-being at work (Koldijk, 2012). Knowledge workers often experience stress building up,

---

<sup>1</sup>R&D project; [www.commit-nl.nl/projects/swell-smart-reasoning-systems-for-well-being-at-work-and-at](http://www.commit-nl.nl/projects/swell-smart-reasoning-systems-for-well-being-at-work-and-at)

which in the worst case results in burn-out (Demerouti, A. B. Bakker, Nachreiner, *et al.*, 2001). The SWELL system collects work related behavior data with sensors, which is interpreted in terms of the user's work context and mental state, to provide personalized feedback and support at work.

New technological possibilities also raise an important question: what do users think about context aware systems that build rich user models? Are they willing to provide their personal data and use this new technology? What are their concerns with these systems? Est *et al.* (2014) wrote a critical book on 'intimate technology', making us aware of potential consequences of new technology that monitors our body and behavior: They state that "the last years, we have given away massive amounts of information about our social life, in exchange for services"; and that "we are now at the point of giving away big amounts of data about our body and well-being." (translated, p. 65). They state that "intimate technology makes us more transparent for the people around us" (p. 28). They ask: "What do we want that others (like parents, the government, insurance companies, police, employers), do and do not know about us?" (p. 65)

As the scale and application of data collection increase, privacy concerns are rising over the new worlds of possibilities. Privacy has been defined by Van De Garde-Perik *et al.* (2008) as the "boundary control process in which individuals regulate when, how, and to what extent information about them is communicated to others" (p. 21). There often seems to be a trade-off for users between using a service and their privacy: they can use a better service by providing more context information, often at the cost of losing control over their personal data. Rubinstein (2012) states that "Businesses inevitably collect and use more and more personal data, and while consumers realize many benefits in exchange, there is little doubt that businesses, not consumers, control the market in personal data with their own interests in mind." (p.1).

To date, there seems to be no compact and coherent approach to designing CAS, that balance functionality and privacy needs. Insights from different research fields need to be brought together to develop a human-centered CAS. Researchers on stress and burn-out may provide useful human factors knowledge, but may know too little about enabling technology. Researchers on sensing and reasoning develop new state-of-the-art technologies, but might be less focused on user needs, concerns and privacy considerations.

Moreover, research in the field of privacy ranges from papers about privacy legislation, over technical security solutions, to user studies on trust. Ideally, important insights on privacy need to be integrated into CAS design. Researchers

found that different type of privacy concerns exist and they change over time (Anton, Earp, and Young, 2010). Smith, Milberg, and Burke (1996) identified data related dimensions of privacy concerns (collection, errors, secondary use, and unauthorized access to information). Anton, Earp, and Young (2010) conducted two surveys in 2002 and 2008 respectively that investigated the level importance of different privacy concerns. Their results showed that the top concerns remained unchanged over time (trading personal data to third parties, desire to be notified about security measures), although the level of concern of specific concerns changed.

It has been argued that respondents that indicate not to be concerned about privacy may not be well acquainted with possible consequences, or may have a false sense of safety due to misconceptions on their IT understanding level (Flinn and Lumsden, 2005). However, even those that do report privacy concerns usually do not act accordingly (Spiekermann, Grossklags, and Berendt (2001), Acquisti and Gross (2006)): a phenomenon known as the privacy paradox. This might be a consequence of poor understanding of possible risks when using a particular service (Flinn and Lumsden, 2005), or of users bounded rationality (Acquisti, 2004): while the risk of information disclosure may be invisible or spread over time (e.g., identity theft), the benefits of disclosing personal information may be immediate (e.g., convenience of placing orders online). Investigating privacy in CAS following a cognitive engineering method, distinguishes our research from related research in which privacy is often investigated in social networks, user profiling, e-commerce, marketing or mobile location enhanced technologies (Smith, Dinev, and Xu, 2011).

**Approach and outline** In this work, we refine the situated Cognitive Engineering (sCE) approach (Neerincx and Lindenberg, 2008) on two aspects: (1) defining the context during the requirements engineering process and (2) addressing functional and non-functional requirements coherently.

The research field of requirements engineering aims to bridge the gap between users and technology, by investigating user needs and translating them into system design (e.g. Sutcliffe (2013)). In scenario-based design, for example, stories about how the technology will be used are formulated ('scenarios'), which promote communication among stakeholders and evoke reflection (J. M. Carroll, 2000). Kaasinen (2003), e.g., used scenario analyses with groups of users to investigate user needs in location aware mobile services, which is similar to the approach we take in our first two user studies. CAS that use context information pose new challenges to requirements engineering. As Bosems and van Sinderen (2014) state: "It is hard to imagine, both for the designer and end-user, all possible relevant contexts and best possible corresponding enriched services."

As a consequence, new approaches become necessary. Seyff *et al.* (2008) propose the development of new tools for requirements engineers to enhance the in-situ requirements elicitation process, e.g. by means of an app that recognizes the current context and provides guidance in specifying relevant requirements for this context. This would, however, mean that relevant contexts are defined up front, which is difficult. Most relevant to our design problem seems to be context-driven requirements analysis (Choi, 2007). This approach helps to first identify specific context aware services that users want. From these identified services, then, the contexts that are necessary to be identified are distilled, and finally, suitable sensors and inference methodologies are determined.

The situated Cognitive Engineering method combines user and operational support demands, with human factors knowledge and possibilities of enabling technology (see Figure 7.1).<sup>2</sup> This results in requirements for the system specification, which are then evaluated and further refined. The main focus is investigating how a system applying user modeling can be designed in a way, that takes the user's privacy concerns into account. In this process, we deploy several complementary analysis, design and test techniques (Note that sCE does not prescribe a specific order in time; the sCE-processes are iterative and can partially take place in parallel).

1. *Addressing human factors knowledge* (Section 7.2.1). Relevant human factors knowledge on stress and interventions is used to identify desired core functions for the CAS.
2. *Assessing enabling technology* (Section 7.2.2). The identified core functions are used to define contexts that the system needs to recognize. By integrating knowledge on technical possibilities, the context attributes and specific sensors are further defined (inspired by Choi (2007)). This results in abstract functional components for the CAS, i.e. its user model, context inference, and the high-level architecture.
3. *Acquiring user and operational support demands* (Section 7.3).
  - In a workshop with potential users, we present the general idea of the CAS, and make an inventory of *user needs*. This results in a list of (mainly functional) requirements.
  - In a second user study, we present a scenario (J. M. Carroll, 2000), the system and the context information it would collect, and make an

---

<sup>2</sup>In this chapter we particularly focus on the user and operational support demands. Our work on stress and intervention theory, and context inference is presented in more detail in Chapters 2 and 6.

inventory of *user concerns*. As a result, we formulate non-functional requirements.

- As *privacy* is an important issue, we perform a Privacy Impact Assessment (Wright, 2012) to identify additional privacy requirements. We then shortly present how 8 Privacy Design Strategies (Hoepman, 2014) can be used to address the privacy concerns.
4. *Harmonizing functional and non-functional requirements* (Section 7.4). We investigate which conflicts exist between our collected functional and non-functional requirements, by means of a Trade-off Analysis (see Sutcliffe (2013)).
  5. *Evaluating for further refinement* (Section 7.5). In a third user study, an implemented prototype is used to investigate opinions on collecting several types of information and using different sensors. In this way, user needs for personalization are identified, as well as corresponding requirements for the user model.

Insights regarding the human-centered development method, the current design rationale of the SWELL ‘CAS for stress reduction’, and the three user studies are presented (Section 7.6). The contribution of our work lies in integrating a variety of methods and techniques for user-centered CAS design, more particularly: analyzing user concerns, complementing the analysis with a privacy impact assessment, and providing techniques to address privacy issues in CAS.

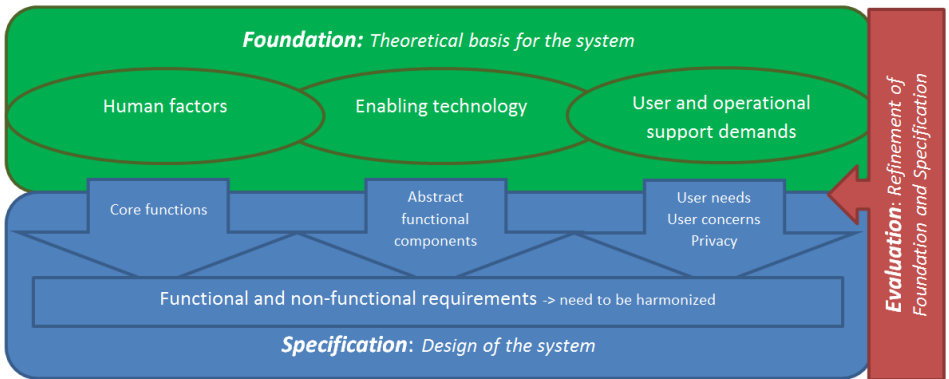


Figure 7.1: Situated Cognitive Engineering (sCE) methodology for user-centered design of privacy-friendly context-aware systems.

## 7.2 Theoretical and technological foundation

We start with describing the theoretical and technological foundation for our CAS for stress reduction.

### 7.2.1 Derive core functions from human factors

The aim of the CAS is to help improve well-being at work. To identify desired context aware services, we use relevant human factors theory on work stress and interventions.<sup>3</sup>

**Human factors theory.** In Figure 7.2 a stress model by Le Fevre, Kolt, and Matheny (2006) is shown. The environment poses a stressor, and depending on characteristics of an individual, this results in a certain perception of stress. As a consequence, stress can be experienced with behavioral, physical and/ or psychological outcomes. In order to alter this stress response, three aspects can be addressed: first of all, the stressor itself could be diminished; second, the individual could cope by altering the perception of stress; and finally, someone can recover from the experience of stress.

In Figure 7.2 we show that the ‘CAS for stress reduction’ could provide 3 context aware services, that can be integrated into the general stress model.

**‘SWELL Workload Mirror’** First of all, insights can be provided in the general stress process. Thus the first context aware service may be described as ‘SWELL Workload Mirror’, which could give feedback on mental effort or stress from which the user can gain insight into his well-being, e.g. prolonged periods of stress. Moreover, the system could give feedback on the relation between the work context and the associated mental effort or stress, from which the user can gain insight on which tasks or topics are demanding and which way of working decreases well-being. Based on these insights, the user could decide to change his manner of working and set goals for himself. The system could also give insight into sleep duration or physical activity, which are relevant to the employee’s recovery.

---

<sup>3</sup>A more elaborate framework on work stress and intervention theory, that was used to develop the SWELL system, is presented in Chapter 2.

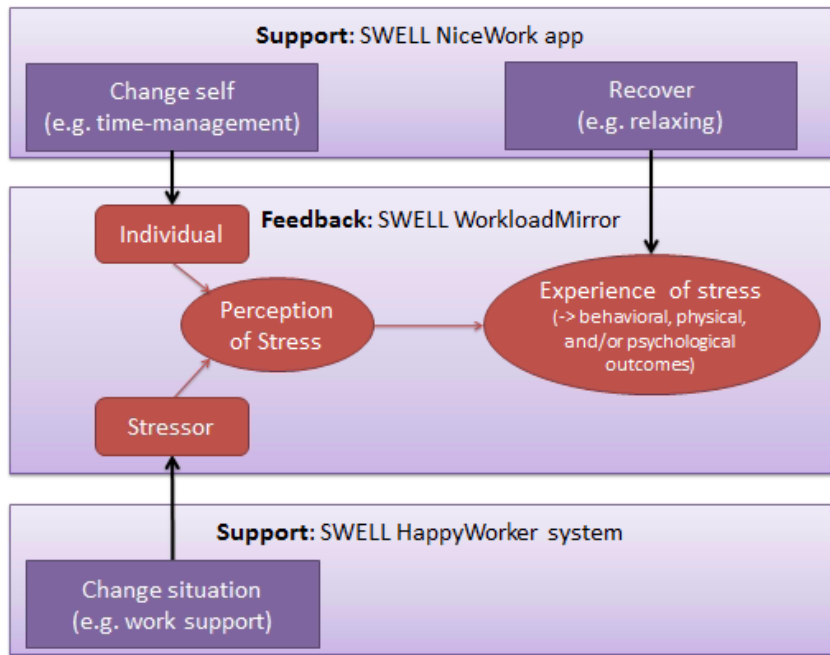


Figure 7.2: Model on stress (Le Fevre, Kolt, and Matheny (2006)) and the associated context aware services for the SWELL system.

**‘SWELL NiceWork’ coaching app** Moreover, the employee himself can be helped, either to change his working behavior or to better recover from stress. Thus the second context aware service may be described as ‘SWELL NiceWork’ coaching app (Wabeke, 2014), which could provide tips to help employees change work behavior, e.g. by giving tips on time-management, or by giving advice on recovery, e.g. relaxing exercises. Based upon information on the stress development over time, the ‘SWELL NiceWork’ app could adapt its feedback, e.g. a tip for relaxation may be most appropriate after an extended period of stress. Moreover, taking the user’s current context into account, the system could provide tips at appropriate moments (e.g. when the user ends a task or when he is at a particular location). Finally, also the content of the provided tips could be determined based upon context information (e.g. suggest a physical activity tip against stress when someone has been very inactive).

**‘SWELL HappyWorker’ system** Finally, the work situation can be changed. Thus the third context aware service may be described as ‘SWELL HappyWorker’ system, that could help to diminish stressors by providing work support. In order to provide effective work support it is crucial to understand the current work context. Based on the current working topic, the ‘SWELL HappyWorker’ support system could for example filter irrelevant email notifications, or provide personalized search.

### 7.2.2 Functional components from enabling technology

After having formulated core functions for the system based on relevant theories, we now consider technological possibilities. The identified core functions are used as starting point to define relevant contexts that the system would need to recognize (inspired by context-driven requirements analysis (Choi, 2007)).

**User model** To provide the desired functionality, the system needs information on the current work portfolio (tasks, topics), as well as on the subjective experience of the work (mental effort/ energy, stress). Moreover, for recovering from stress, private aspects outside of work may also be interesting to include in the user model (amount of sleep, physical activity, social interactions, the current location). This information is represented in the system’s user model (see Figure 7.3).

**Context inference** Ideally, the context information is automatically inferred from sensor data. Following the methodology by Choi (2007) we first identify relevant context attributes and then specific sensors that can be used (see Figure 7.4). In the SWELL project we focus on unobtrusive, relatively cheap sensors that can easily be used in office environments.

Regarding the recognition of *work context*, we would like to recognize the current task and topic someone works on. We developed algorithms to infer *work task* (e.g. write report, search information) from several computer interaction features (Koldijk, van Staaldin, Neerincx, *et al.*, 2012). Moreover, Sappelli (2016) developed algorithms to infer *work topics* from computer interactions and content information from the user’s computer. To capture relevant data, a computer logger can be used.

As inference of *well-being at work* is a new domain, we did a literature study on the technical possibilities of inferring stress and related aspects from sensors:



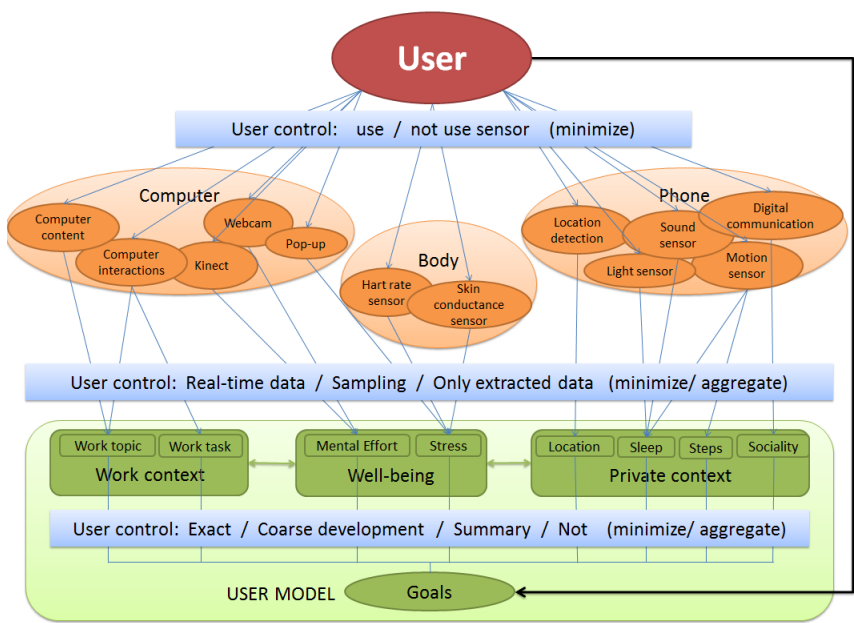


Figure 7.3: Overview of the ‘CAS for stress reduction’ and its user model, which holds information on the users work context, and his well-being. Moreover, information on the private context outside of work may be included. Most important in the design is the aspect of user control (explained later).

**Enabling technology** With respect to inferring stress from sensor data, Sharma and Gedeon (2012) provide a compact survey. Often, body sensors are used to measure the physiological stress response directly, e.g. skin conductance (J. Bakker *et al.*, 2012) or heart rate (Hogervorst, Brouwer, and W. K. Vos, 2013). These can, however, be too obtrusive for the user who would have to wear sensors on his body. There also is potential in using outward characteristics, such as facial expressions, postures or computer interactions as indicators for the user’s mental state. Facial expressions are currently mainly used for inferring emotions, but facial expressions could also show cues to infer other mental states that might be more relevant in a working context. In a study where working conditions were manipulated with stressors, we found that specific facial action units may be indicative

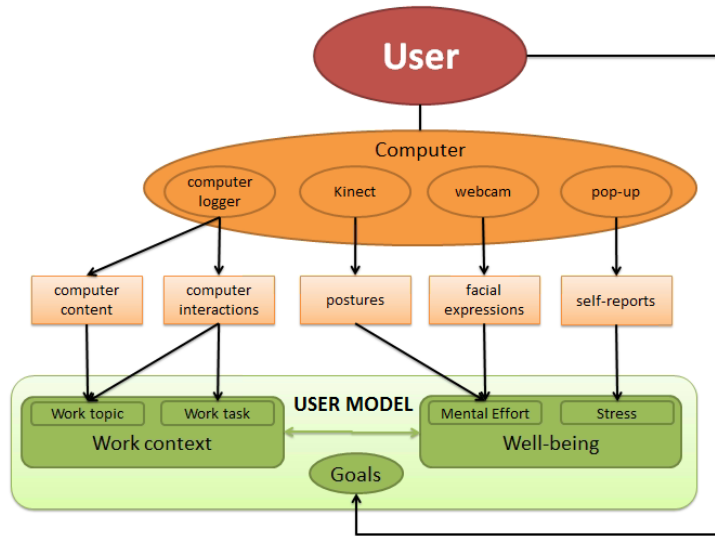


Figure 7.4: Inference of context information for the user model.

of experienced mental effort (Koldijk, Sappelli, Neerincx, *et al.*, 2013). Research by Dinges *et al.* (2005) suggest that facial activity in mouth and eyebrow regions could be used to detect stress. Moreover, Craig *et al.* (2008) looked at facial expressions while students worked with an online tutoring system. Association rule mining identified that frustration and confusion were associated with specific facial activity. Mental states are also being estimated from computer interaction data. Results by Vizer, Zhou, and Sears (2009) indicate that stress can produce changes in typing patterns. Finally, Kapoor and Picard (2005) describe work on recognizing interest in students by means of computer interaction and posture.

Our work on inferring the user's mental state from sensor data is ongoing (see Koldijk, Sappelli, Neerincx, *et al.* (2013)). To infer the experienced *mental effort*, we are investigating the use of facial expression data. To capture facial expressions, a camera can be used. Additionally, we are investigating the use of posture information. To capture postures, a Kinect 3D camera can be used. Moreover, to capture stress in the user model, a pop-up may be used to ask the user for self-reports. We are also investigating the use of physiology (heart rate,

skin conductance). To capture physiology body sensors can be used.

Finally, personalization plays an important role. General models will have to be adapted to the specific user to reliably estimate work tasks, topics or mental effort.

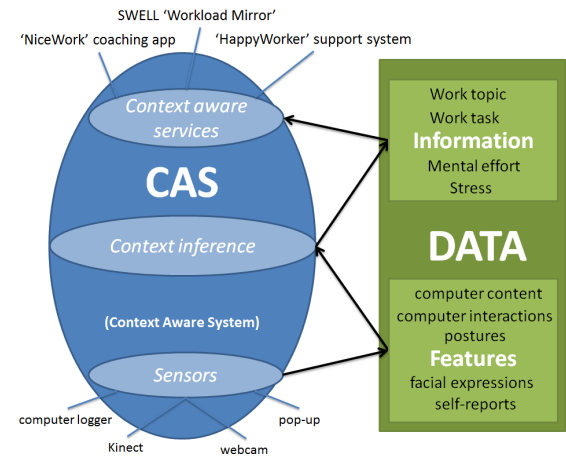


Figure 7.5: High-level architecture. The ‘CAS for stress reduction’ has a layer with various sensors, from which context attributes are extracted and stored. The context inference layer uses the context attributes to infer contexts. This context information is stored in the user model and can be used by several context aware services (terminology according to Choi, 2007).

**High-level architecture** We now describe the SWELL system as a Context Aware System, see Figure 7.5. It consists of a layer of sensors, a layer of reasoning and an application layer. Next to the CAS system, there is a data layer. The SWELL system uses various sorts of sensors, e.g. computer logging or a webcam, to collect so called primary data (Perera *et al.*, 2014). From the sensor data useful features are extracted, e.g. computer activity or facial expressions. This data is stored and can be used by the reasoning component to infer higher level information or so called secondary data (Perera *et al.*, 2014), e.g. the tasks performed or mental effort. According to Baldauf, Dustdar, and Rosenberg (2007) and Prekop and Burnett (2003) this would be referred to as “internal context data”. The context information about the user is stored in the user model and can be used by an application to provide its context aware service, e.g. presenting the user the ‘SWELL WorkloadMirror’. Other context aware services

can also use the available context information. The ‘SWELL NiceWork’ coaching app could use context information to offer tips to the users. This would be described by Perera *et al.* (2014) as passive context awareness. The ‘SWELL HappyWorker’ support system could use context information to take action for the user, e.g. hold back irrelevant emails. This would be described by Perera *et al.* (2014) as active context awareness.

**Conclusions on the methodology** By applying context-driven requirements analysis (Choi, 2007), we were able to first focus on specifying desired context aware services. Then, we identified the underlying context information that would be necessary to provide the services. As a last step, we determined underlying types of information that would be necessary for context inference, and selected appropriate sensors and algorithms. In this way the functionality of the system is leading. The developer then applies his domain knowledge and expertise in context inference to decide upon the specifics of the underlying system. We think this is in general a very suitable approach for the design of context aware systems.

## 7.3 User and operational support demands

After having described the theoretical and technological foundation, we now describe the user and operational support demands. We identify user needs, user concerns, and specifically address privacy.

### 7.3.1 Identify user needs

To get a clear picture on the specific context aware services for the ‘CAS for stress reduction’, we not only considered literature on stress and interventions, but also performed a user study on ‘User needs’.

#### Method (user study 1)

**Design** A two hour workshop on user needs was organized. The following aspects were addressed: 1) What should the CAS for stress reduction be able to do? 2) When and how is the system used? 3) Which aspects have to be taken into account when introducing the system? The first question was used to gather functional requirements, the second to gather requirements on the context of use, and the third to identify barriers. The input of the participants was analyzed qualitatively. We structured and summarized all user input and derived requirements.

**Procedure** The workshop started with an introduction in which we shortly explained that the system should support self-management of well-being at work, which means that it should help knowledge workers to observe, judge and change their behavior. For each of the questions described in the Design, a 30 minute time-slot was allocated. For each question, we first presented some ideas from interviews with knowledge workers who had previously dropped out of work because of burn-out (2 female, 3 male; aged between 30 and 50; Vos, 2011). This input was used as starting point for the participants in the workshop. They were then asked to write down their additional ideas individually, imagining the optimal, perfect system, without thinking about technical possibilities and limitations. The sensor that we envisioned to use was at this stage only computer logging. Finally, all ideas were discussed in the group. During the sessions we took notes on the users' input.

**Participants** Seven people (all knowledge workers, with a technical background; 1 female, 6 male; aged between 25 and 45) participated. The study was run at TNO (Netherlands Organization for Applied Scientific Research), with a staff mainly consisting of researchers, consultants, project leaders and developers. We asked interested colleagues to participate.

## Results

Users indicated that the system should provide an overview of performed work, preferably in combination with work behavior and the associated subjective experience. This information can then be used by the user to gain insight in work processes. The collected information can also be used to provide services to the user: Information on the user's context can be used to determine the content and timing of support. Important is the flexibility of the system, e.g. regarding the sensors used and the frequency of feedback. We also identified some important factors to address, e.g. not irritating users and addressing privacy.

## Conclusions

The derived functional requirements are presented in Table 7.1, and some identified non-functional requirements are included in Table 7.2.

### 7.3.2 Identify user concerns

To further develop a user-friendly context aware system, we set up a second user study to investigate potential hurdles to use the 'CAS for stress reduction'.

---

RF01:	The system shall give insight in spent time, in terms of activities over the day and topics worked on.
RF02:	The system shall provide awareness on working behavior.
RF03:	The system shall enable the user to set personal goals.
RF04:	The system shall enable good work planning.
RF05:	The system shall help finding causes of good/ bad work-days.
RF06:	The system shall signal deviations and trends.
RF07:	The system shall (optionally) support behavior change by giving advice.
RF08:	The system shall (optionally) provide norms.
RF09:	The system shall (optionally) actively interfere to facilitate a better way of working.
RF10:	The system shall provide support at appropriate moments.
RF11:	The user shall be able to personally adjust which aspects to monitor.
RF12:	The user shall be able to personally adjust how to visualize data.
RF13:	The user shall be able to personally adjust how often the system gives feedback.
RF14:	The system shall also provide positive feedback.
RF15:	The system shall use a funny and playful manner, e.g. a game element.

---

Table 7.1: Functional requirements for the CAS for stress reduction, resulting from user-study 1 (User needs).

### Method (user study 2)

**Design** A one hour session was organized in which Scenario-based Design (J. M. Carroll, 2000) was used to identify user concerns with the CAS for stress reduction. The input of the participants was analyzed qualitatively. The concerns of all participants were manually clustered into groups of similar concerns. For each type of concern we counted how many participants mentioned this concern.

**Procedure** The session started with a presentation of a scenario (see also Figure 7.6). Then, the participants were asked to write down all their potential concerns with the system. Their notes were finally collected for further analysis.

*SWELL Scenario* “Bob is 40 years old and works in an office from 9 to 5, where he performs knowledge work. Since some time now, Bob feels some tension and finds it hard to get work off his mind in the evenings. At the end of his working day he often notices that he has not completed all planned tasks and he feels stressed. Bob decides to use the SWELL system. At the end of his working day he opens the ‘SWELL Workload Mirror’ to look back at his day. He sees an overview of the tasks he performed and content he worked on, combined with information on his subjective energy level. He sees that he worked very fragmented and notices that this probably caused his loss of overview and decline in energy. Bob decides that it would be better for him to stay focused on his planned work and determine a time-slot to do all ad-hoc tasks. He enables a functionality of the SWELL system, which warns him when he makes too many

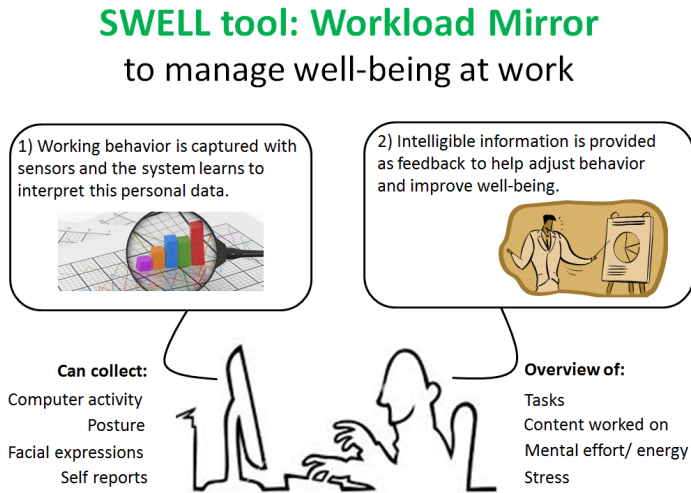


Figure 7.6: Information about the SWELL system.

task switches again. Bob also notices that, in fact, he has done a lot of useful things today and can go home satisfied.”

**Participants** The session was held during a monthly department meeting at TNO. All present colleagues were asked to participate. All eleven colleagues participated (knowledge workers with a technical background; 2 female, 9 male; aged between 25 and 45). <sup>4</sup>

**Results**

Many participants (7 out of 11) had concerns about who could access their data. Many explicitly mentioned that they would not want the data to be shared with the management. Many participants were also afraid that the system would require effort, which might not outweigh its benefits. About half of the participants (6 of 11) mentioned that they would want to know exactly what happens with the data, e.g. what is stored and where. About half of the participants (5 of 11) had concerns regarding the performance of the system, e.g. slowing down the computer. About a third of the participants (4 of 11) had doubts about

<sup>4</sup>The disadvantage of asking colleagues was that the responses may have been somewhat biased. The advantage, however, was that people from this department were very familiar with the aspect of data processing in general and could provide valuable input.

the system, in the sense that they were not sure whether it would solve their problem. About a fourth of the participants (3 of 11) had issues regarding the quality, e.g. the reliability of the inferences made by the system. About a fourth mentioned issues regarding their control over the system. Finally, 2 participants mentioned that they would feel monitored.

### **Conclusions and discussion**

From all mentioned user concerns we extracted additional non-functional requirements for our ‘CAS for stress reduction’, which can be found in Table 7.2.

Our results on user concerns are comparable to what Knijnenburg and Kobsa (2013b) (see online Appendix) found in their study on user concerns regarding information disclosure in context-aware recommender systems. In line with our finding, they describe that many users were concerned of other uses of the provided information, e.g. for surveillance. However, in their case, users thought that advertisement was the most likely unintended use. Moreover, they also found that users did not want to answer requests that took too much effort. Incorrect inference was also a concern. Contrary to our findings, the users were less interested in what exactly happens with the data, e.g. what is stored and where. This difference may be explained by the fact that our participants had a rather technical background.

Based on the users’ concerns, some balances in the system design need to be found. Concerns regarding the effort the system would require, highlight the importance of automatic inferences and smart support, while users also wish to stay in control. A human-in-the-loop approach, combining automatic processing with human interaction seems a good solution. Furthermore, inferences should be reliable, while not slowing the computer down. Solutions can be using CPU and memory efficient algorithms, running inference algorithms on a server or just analyzing samples of data. We will elaborate on these trade-offs in Section 7.4 and will further address them in our user study presented in Section 7.5. Moreover, we found that many concerns are related to privacy, i.e. who can see the data, what will happen with the data, sharing data with the management and the feeling of being watched. We will elaborate on privacy in the next sections. Finally, some users express doubts about the ‘CAS for stress reduction’, so this solution to more well-being at work may not be suitable for every user.

#### **7.3.3 Address privacy**

As privacy is an important aspect in CAS that build rich user models, we particularly focus on privacy requirements. We performed a privacy impact assessment



and outline how privacy by design can be used to address privacy concerns.

### Privacy Impact Assessment

In order to get a better insight in privacy aspects of our ‘CAS for stress reduction’ we performed a Privacy Impact Assessment (PIA). As Wright (2012) describes it: “PIAs provide a way to detect potential privacy problems, take precautions and build tailored safeguards before, not after, the organization makes heavy investments in the development of a new technology, service or product.” (p. 54). For more information on PIAs, see e.g. the UK and New Zealand PIA Handbooks (Office (2014); Commissioner (2007)). We used the PIA question catalog by Norea<sup>5</sup>. This catalog provides a structured manner to analyze potential privacy risks before implementing a product, service or proposed legislation. As a consequence, the question catalog is very elaborate and has a broad focus, including many questions not relevant to the design of CAS. Moreover, it lacks important questions regarding (sensor) data, inferences and user models. We structured the questions and extracted the topics that were most relevant to the design of CAS. Details are reported in Koldijk, Koot, *et al.* (2014)<sup>6</sup>.

The most important privacy considerations for CAS that we identified are: 1) Goal of data collection - Only when users understand what the system does and why the collection of data is necessary, they will be able to take a well informed decision on how to use the system; 2) Type of data - No more data than necessary should be collected and data should be stored in the most aggregated form that is still useful; 3) Data sharing - It is important that the user gives consent and that the data is used in the intended way; 4) Reactions to the system - Reactions to new innovative systems are hard to predict; 5) User control - The user should be in control of the system and the settings; 6) Quality of the data - The data should be up-to-date, correct and complete; 7) Security of the data - Unwanted or unauthorized access of the data should be prevented; 8) Data responsibilities - The more parties are involved, the higher the risk of data getting lost, unclear responsibilities or use of data for other purposes.

We used these PIA results to formulate additional (privacy and security) requirements for the ‘CAS for stress reduction’, see Table 7.2. Note that several aspects that we identified with the PIA were not yet mentioned by users themselves in the previous user studies.

**Conclusions on the methodology** We recommend applying a PIA, besides running user studies, to formulate a complete list of requirements. Being aware

---

<sup>5</sup><http://www.norea.nl/Norea/Actueel/Nieuws/Presentatie+PIA.aspx>

<sup>6</sup>Chapter 8

of these requirements at an early stage of design enables developers to implement privacy into their context aware system. This aids the development of more privacy-friendly CAS. In the next section we provide more detail on Privacy by Design.

### Privacy by Design

In this section we present how the outlined privacy aspects can be addressed from the developer's perspective by using Privacy by Design. Privacy by Design (PbD) seeks to embed privacy and data protection into the design specifications of information technologies, business practices and networked infrastructures (Cavoukian, 2009). Privacy cannot be achieved by implementing some add-on, but is an integral part of the entire system and should be taken into account during design. The new European Data Protection Act will be more strict than the current version and will require applications to be more privacy friendly. Furthermore, we argue that a privacy-friendly system increases end-user trust, which results in higher uptake of the service, generating a positive business case for a privacy-friendly version of the CAS.

Hoepman (2014) describes eight Privacy Design Strategies that can be used in early stages of software development to design a privacy-friendly system. The strategies are derived from Solove's Taxonomy of Privacy (Solove, 2006) and the EU data protection legislation (EU, 1995), and mapped on the ISO Privacy Principles. The strategies are simple, easy to apply and very helpful in making good design decisions in order to develop a context-aware system that follows current privacy legislation. Details are reported in Koldijk, Koot, *et al.* (2014)<sup>7</sup>. Here we shortly summarize how each of the strategies can be applied to address the privacy requirements in the SWELL case:

- Inform - It is important to inform the user about which information is processed, for what purpose, and by which means. In SWELL, we focus in particular on transparency by informing users. Users should know in detail which data is collected, what information is stored in the user model, where it is stored and for which aim the data is used. The purpose limitation should be stressed: The information in the user model is only used for helping the user to reach his well-being goals.
- Control - Moreover, it is important to give the user control over the data and what can be done with it. Always Informed Consent (Romanosky *et al.*, 2006) should be used, which means getting permission from the user

---

<sup>7</sup>Chapter 8

to collect data for a specified purpose. In SWELL, the user will have full control over the data, can view or delete it. Moreover, the user will be able to enable or disable every sensor. We elaborate on the aspect of control in Section 7.5, where we present a user study.

- **Minimize** - The task of the designer of the system is to minimize the amount of data that the system stores. The SWELL system will only processes data that is necessary to provide the functionality that is desired. We address user preferences on data minimization in our user study presented in Section 7.5.
- **Aggregate** - By applying reasoning, the data can often be aggregated even further. In SWELL, the sensor data will be processed locally on the users' device. Only summary information, such as topics, average posture or facial expression, will be stored – no keystrokes or video. We address user preferences on data aggregation in our user study presented in Section 7.5.
- **Hide** - The developer should take care to hide the information, such that the data strictly belongs to the user and cannot be accessed unauthorized. To ensure the security of the data it is a good idea to store data encrypted. In SWELL the data will be hidden from unauthorized access. However, data security is not the focus of this work.
- **Separate** - By separating the processing or storage of several sources of personal information that belong to the same person, no complete profiles of one person can be made. Data should be processed locally whenever possible, and stored locally if feasible as well. One of the main aims of SWELL is combining different sorts of data into one user model, to provide users insight. We envisioned the use of one central place to store and combine all data. We may want to rethink this architecture and apply data separation. Moreover, storing data locally may have advantages regarding privacy, but also disadvantages for the functionality of the system (e.g. slowing down the device, data not accessible from other locations). We elaborate on such trade-offs in Section 7.4 and also address them in our user study in Section 7.5.
- **Enforce and demonstrate** - Enforcement of a privacy policy compatible with legal requirements is necessary. Moreover, an organization must be able to demonstrate compliance with its privacy policy. It is important in SWELL to enforce and demonstrate that the system fulfills current legislation around privacy, however, this is not the focus of this work.

In the overview of requirements for the ‘CAS for stress reduction’ (Table 7.2) we added information on which Privacy Design Strategy is used to address each specific requirement.

**Conclusions on the methodology** By applying all 8 Privacy Design Strategies, the CAS can be designed such that privacy concerns are handled in a suitable manner. A paper related to our work is the survey by Toch, Wang, and Cranor (2012) on privacy challenges and technologies to reduce privacy risks. They focus on social-based personalization (e.g. in social networks), behavioral profiling (e.g. for advertising) and location-based personalization. These systems pose different privacy challenges than a CAS for stress reduction does, as described earlier. The technologies described to address these issues are however similar to the ones described here: the use of pseudonyms, storing user profile data locally on the users’ device, aggregation and obfuscation of data, user control and feedback. This highlights their broad applicability.

In Figure 7.7 we integrated the eight privacy design patterns into the CAS view. We find that of the 8 general Privacy Design Strategies described by Hoepman (2014) the strategies ‘minimize’, ‘aggregate’ and ‘control’ are particularly interesting in the view of CAS. When connecting sensors, we cannot only think about minimizing the amount of sensors to be used, but also about aggregating the raw data into a format that is less privacy sensitive. By means of pattern recognition and machine learning we can infer higher level information without the need to store low level data. Processing data on the fly, however, poses interesting new technical challenges, for example: do we apply machine learning locally on the user’s device which may be most privacy friendly, or do we process all data in the cloud which may be more efficient. The principle of ‘control’ suggests that users should be able to make choices on which data is collected and how this data is used. We elaborate on this in Section 7.5. So designing CAS in light of privacy by design poses as well new solutions as new challenges that should be investigated.

We have to note that the final implementation of privacy in the system can be difficult. Incorporating privacy may require a different architecture, which is more complicated than the one simply enabling all desired functionality. This may be an extra burden to developers. Moreover, additional privacy requirements can be contradictory to general functional requirements. We will elaborate on the trade-offs that have to be taken in Section 7.4.

	<i>Non-functional requirements</i>	U1	U2	PIA	PD
System	RSy20: The system should be easy to install.	X			-
	RSy21: The system should work on all work locations computers.	X			-
	RSy22: The system should not slow the computer down.		X		-
Usability	RU23: The system should be easy to use.	X	X		-
	RU24: The system should not be time consuming.	X	X		-
	RU25: The system should not provide too many pop-ups/ interventions.	X	X		-
	RU26: The user should be able to turn the system on or off.	X	X		-
	RU27: The user should be in control of the system.	X	X	X	-
	RU28: The user should be in control of the settings (e.g. personally adjust how often the system gives feedback).	X		X	-
Quality	RQ29: The system should give correct information.	X	X	X	-
	RQ30: The user should be able to check and correct the data.			X	C
Privacy	RP31: The user's privacy concerns should be adequately addressed: The system shall process personal data for which the purpose is specific, clearly defined and legitimate.	X	X	X	E
	RP32: The user should have a clear view on what the system does and how the data is used (transparency).	X	X	X	I
	RP33: The user should have the possibility to delete data.	X	X	X	C
	RP34: The data should be used for individual purposes only.	X	X		C
	RP35: The data should not be shared with the manager.	X	X		C
	RP36: Others should not have access to your data.		X	X	C
	RP37: The goal of data collection should be clearly described.			X	I,E
	RP38: The user must know which data is collected.			X	I
	RP39: The type of collected sensor data should be suitable to fulfill the desired goal.			X	A
	RP40: The user must give permission to collect data, based on a well-informed decision.			X	I,C
	RP41: Which data is collected and processed will be kept to a minimum to enable required functionality.			X	M
	RP42: The system should provide an alternative means to provide data (e.g. manual user input, in case a user does not want to use a sensor).			X	-
	RP43: The user should be able to see his own data.			X	C
	RP44: A clear data description should be made.			X	E
	RP45: The user may decide to share the data (user in full control of personal information sharing).			X	C
	RP46: When the user voluntarily shares data, it should be shared in line with the user's expectations.			X	C
	RP47: The user must know who (if applicable) will have access to the data			X	C
	RP48: The user should be aware of his privacy settings.			X	I
Security	RS49: The data should be encrypted.		X	X	H
	RS50: The data should be stored as locally as possible.		X	X	S
	RS51: The data should be stored as aggregated as possible.			X	A
	RS52: It should be prevented that different sorts of data can be combined to yield privacy sensitive conclusions.			X	S
	RS53: The system should not store identifiers as full names and email addresses.			X	M
	RS54: The system shall not keep personal data in a form which allows users to be identified for any longer than necessary.			X	M
	RS5: An security plan should be established to prevent unauthorized access.			X	H
	RS6: All involved parties should adhere to the security plan.			X	-

Table 7.2: Overview of non-functional requirements and the stage in which they were identified: user-studies (U1, U2) and privacy impact assessment (PIA). Note that the PIA helped us to add many additional (privacy and security) requirements, not yet mentioned by the users. The last column denotes how the requirements can be addressed with Privacy by Design (PD) and by which strategy: Inform, Control, Minimize, Aggregate, Hide, Separate, Enforce&Demonstrate.

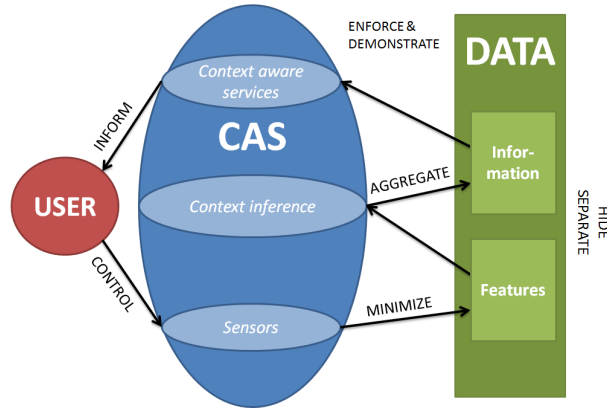


Figure 7.7: Privacy Design Strategies in the CAS view.

## 7.4 Harmonizing all identified requirements

In the first part of this chapter we identified core functions and functional components for the ‘CAS for stress reduction’. We also identified a set of requirements, based upon user needs, user concerns and privacy. As noted in the introduction of this chapter, there is often a trade-off for users between using a service and their privacy. Implementing privacy by design is useful to take away some of the privacy risks. However, some of the privacy and security requirements can be conflicting with requirements on the system’s functionality and usability.

To investigate potential conflicts in our identified requirements, we performed a Trade-off Analysis (see Sutcliffe (2013)). Some examples of conflicting requirements are given in Figure 7.8. The security requirement of preventing the combination of different sorts of data (RS52) is in conflict with the functional requirement of providing insight in working behavior (RF02): this insight might be especially powerful when different sorts of data are presented together (e.g. this content relates to this stress level). Another example of conflicting requirements is, where to process and store data. Processing and storing data locally is beneficial for security (RS50). Processing and storing data on a server means less processing demands for the device (RSy22) and data is available for the user from different locations or devices (RSy21). Moreover, the usability requirements of being easy to use (RU23), not time consuming (RU24) and providing not many pop-ups (RU25) can be contradictory to the usability requirements of being in control of the system (RU27) and the settings (RU28): controlling the

system and settings may require additional effort and time from the user. To implement the system, these conflicting requirements need to be harmonized. In the user study presented in Section 7.5 we asked users to make a choice for several of these trade-offs.

Requirements	RS52: It should be prevented that different sorts of data can be combined.	RS50: The data should be stored as locally as possible.	RS51: The data should be stored as aggregated as possible.	RS49: The data should be encrypted.	RU28: The user should be in control of the settings.	RU27: The user should be in control of the system.	RP45: The user may decide to share the data.
RF02: The system should provide insights into working behavior.	X						
RSy21: The system should work on all work locations/ computers.		X					
RSy22: The system should not slow the computer down.		X	X	X			
RU23: The system should be easy to use.					X		
RU24: The system should not be time consuming.					X	X	
RU25: The system should not provide too many pop-ups/ interventions.					X	X	
RP34: The data should be used for individual purposes only.							X

Figure 7.8: Conflicting requirements. Often trade-offs between desired functionality of the system (rows) and privacy and security aspects (columns) play a role.

Besides collecting data for own use, it might also be interesting for the user to share data with other people in specific contexts. Within the SWELL project, work on privacy control has been done by Hulsebosch *et al.* (2013). They present and evaluate mock-ups, in which the user can select which data will when be shared with whom and for what reason. They suggest that context information could be used, to automatically reason about the applicable privacy policy, making privacy control less intrusive for the user. Besides a privacy trade-off at an individual level, there can also be a trade-off at a group or team level: Sharing information can both violate individual's team-member privacy and improve team performance (e.g., by conveying cognitive and affective limits in team-member's task performance and opportunities for reallocating his or her tasks). For critical problems, formalization and simulation can help to establish a well-considered decision. Harbers, Aydogan, *et al.* (2014) provide such an approach for the type

and regularity of information sharing between train traffic controllers. This approach complements the privacy by design and requirements analysis presented here, and this combination could bring about sound privacy-sensitive, individual and group, context-aware support systems.

**Conclusions on the methodology** In general, privacy and functionality requirements can cause conflicts. Taking a design decision may entail the weighing of several trade-offs, for which it is hard to decide what the best choice is. Knowing the existing trade-offs is a first important step. In further system design such trade-offs may need to be settled, or, when possible, the user may be left the choice to make this trade-off for him- or herself (see Section 7.5). This could be the decision to either use or not use the entire system. More elegantly, the settings should be flexible enough to accommodate for different users' (privacy) preferences (Bokhove *et al.*, 2012).

## 7.5 Evaluation for further refinement

Now, we have a first full specification of the CAS for stress reduction, and are aware of conflicting requirements. As a final step, we performed a formative evaluation with a small set of participants who actually used the 'SWELL NiceWork' e-coaching app (Wabeke, 2014), for further refinement of the system. The focus is on aspects of user choice and control regarding the collection of data and the use of inferred context information for supporting well-being at work. We differentiate several types of information and sensors, as was done in Klasnja *et al.* (2009). We focus on the subjective perception of information sensitivity, together with information usage (Adams, 2000).

### 7.5.1 Method (user study 3)

**Procedure** All participants had previously tested the 'SWELL NiceWork' coaching app for two weeks. So they had a feel for an app that provides coaching throughout the workday. This app was not yet context aware, so we invited the participants to test a second smartphone app developed within the SWELL project: the CommonSense Tracker app (see Materials). By using this tracking app for a few days, the participants were able to get an impression and feel for the possibilities of context and activity tracking, since they had access to a personal dashboard which provided various views on the collected user model data. Most participants tested the CommonSense Tracker app for 4-8 days, one participant had used it three weeks and one even one year (started using the app prior to the user study). Participant 'a' had no time to test the CommonSense



Tracker app. In a questionnaire, they were then asked to set the configurations for data collection to be used for own insight and for improving the coaching app.

**Materials** ‘SWELL NiceWork’ coaching app (Wabeke, 2014). – The NiceWork app provides well-being tips to knowledge workers (see Figure 7.9). The system learns from user feedback and adapts the content of the tips to the preferences of the user (by using a recommendation engine). In a previous study we had found that users appreciated the app and followed up 2 (of 3) tips per day (Wabeke, 2014). However, it was found that in many cases where a tip was not followed up, this was due to inappropriate timing. Context awareness could thus improve the NiceWork e-coach.

**CommonSense Tracker app**<sup>8</sup> – The CommonSense tracker app uses sensors in the smartphone to infer the user’s context in terms of: location, steps, time active, sleep and sociality (see Figure 7.9). The raw data is condensed and interpreted by the app, and uploaded to the personal cloud based data platform. A dashboard is provided, in which the user can see his own data.

**Questionnaire.** The questionnaire had items on: which information would users want to collect for their own insight (in what detail and why); which information they would like to share with the SWELL coaching app (in what detail and why); which sensors may be used to collect data (in what detail and why); how privacy sensitive the user finds each type of information and each sensor and what they are afraid of; and a set of items in which users were asked to select their preference out of two conflicting requirements.

**Design** The main focus of our user study was: which information do users want to collect in a user model (for own insight and for improving the coaching app). Moreover, we asked in what amount of detail they would allow data collection per sensor. The study had a qualitative nature and we asked the users to explain their choices. We also explicitly asked how privacy sensitive users found different sorts of information and different sensor recordings.

We investigated the following types of information: 1) The types already included in the tracking app: location, sleep, steps/ time active and sociality; 2) Other types of information relevant to the user model: work tasks (e.g. write report, email), work topics (e.g. project A, B, C), manner of working (e.g. derived from body posture or facial expressions), mental workload and stress.

We investigated the following sensors: 1) The sensors already included in the tracking app: motion sensors (accelerometer, gyroscope etc.), location detection

---

<sup>8</sup><https://www.commonsense-dashboard.com>. Available in the Google play store: [https://play.google.com/store/apps/details?id=nl.sense\\_os.smellyfeet&hl=en](https://play.google.com/store/apps/details?id=nl.sense_os.smellyfeet&hl=en)

(GPS, cell network antenna etc.), light sensor, sound sensor and digital communication (calling, text messages etc.); 2) Other types of sensors to enrich the user model: computer interactions (clicks, application changes etc.), computer content (websites, documents), web-cam (facial expressions), Kinect 3D camera (body posture), heart rate sensor and skin conductance sensor (for estimate of stress).

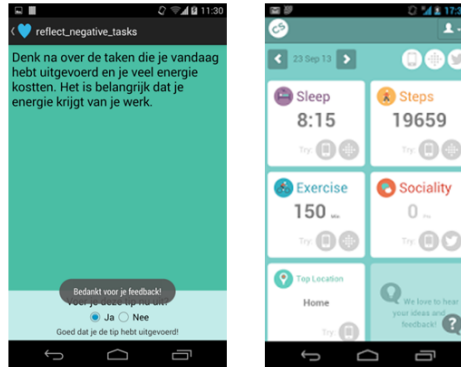


Figure 7.9: Left: NiceWork app, providing a well-being tip (and asking for user feedback). Right: CommonSense Tracker app, showing the dashboard with collected information on sleep, steps, exercise, sociality and location.

**Participants** Participants from a previous user study with the NiceWork e-coach app were emailed to take part in a short follow-up study. Nine people accepted to participate in this user study (aged between 30 and 50, 4 female, all TNO employees). They received a subject fee. Most participants indicated that they would want help to improve their well-being. The extent of general privacy concerns was mixed among participants.

## 7.5.2 Results

First of all, we asked for reasons to not use the CommonSense Tracker app. Only one user mentioned privacy (uncertainty about ownership of the data). Users were more concerned about battery use, or having no interest in the collected data.

In general, there is a large difference between users, on which information they would want to collect in a user model and in what amount of detail (exact development over the day, coarse development over the day, or only summary

per day), see Table 7.3. We asked users to explain their choices. We found that the type of information that users want to collect for own insight depends on their personal interests and goals (e.g. regarding behavioral change). The amount of detail depends on the benefit they see.

Type of information	exact develop- ment over the day	coarse develop- ment over the day	only a sum- mary per day	not
location	2 (c,g)	2 (a,i)	2 (d,h)	3 (b,e,f)
sleep	0	3 (d,h,i)	5 (a,b,c,e,g)	1 (f)
steps/ time active	5 (b,d,g,h,i)	2 (a,c)	1 (e)	1 (f)
sociality	1 (h)	4 (a,c,e,i)	1 (d)	3 (b,f,g)
work tasks	3 (c,e,h)	4 (b,d,f,g)	2 (a,i)	0
work topics	3 (c,e,h)	4 (b,d,f,i)	2 (a,g)	0
manner of working	3 (c,g,h)	3 (b,e,f)	2 (d,i)	1 (a)
mental workload	4 (c,f,h,i)	4 (b,d,e,g)	0	1 (a)
stress	5 (c,d,f,h,i)	2 (b,g)	1 (e)	1 (a)

Table 7.3: Which information do users want to collect for own insight and in what detail. (Amount of users that chose the option denoted, together with user ids.)

In general, there is a difference between users, in which information they consider privacy sensitive (see Table 7.4). As slight trend we see that mental workload and stress is often perceived as privacy sensitive. On the other side, steps/time active is often perceived as less privacy sensitive. Users are mainly concerned of misuse of the data by colleagues, the boss or externals (e.g. insurance companies).

Type of information	very sensitive	a little sensi- tive	neutral	not very sensi- tive	not sensi- tive
location	4 (b,c,e,f)	2 (d,h)	0	1 (i)	0
sleep	2 (g,h)	2 (d,c)	3 (b,e,i)	1 (f)	0
steps/ time active	0	0	2 (a,h)	6 (b,c,d,e,f,i)	0
sociality	3 (b,g,i)	1 (c)	3 (d,e,h)	1 (f)	0
work tasks	2 (b,i)	4 (c,d,e,h)	1 (g)	1 (f)	0
work topics	3 (b,e,h)	3 (c,f,i)	1 (d)	0	0
manner of working	3 (a,b,g)	3 (c,i,h)	1 (d)	2 (e,f)	0
mental workload	6 (a,b,d,g,h,i)	1 (c)	1 (e)	1 (f)	0
stress	6 (a,b,d,g,h,i)	2 (c,e)	0	1 (f)	0

Table 7.4: How privacy sensitive the users consider each type of information. (Note: due to a technical issue responses for participants ‘a’ and ‘g’ on ‘a little sensitive’ and ‘not very sensitive’ were not properly recorded.)

In the current set-up the collected user model data is only available for own insight in the CommonSense tracker app. We asked the participants whether they would want to share the collected information for use in the NiceWork e-coach app. We explained per type of information how this could improve the content and timing of provided tips. All participants, except participant

‘a’ (who had not used the tracking app) wanted to connect the CommonSense Tracker app to the NiceWork app. In general, there is again much difference between users, which information they would want to share with the NiceWork e-coach app. Again, we asked users to explain their choices. We identified several different strategies. Some people stated that they want to share everything they collect with the e-coach. Their rationale behind this is that privacy should be warranted and then all data would be safe. Other people stated that they would only want to share data which is relevant to their personal goal, or that is necessary for the e-coach to work well.

In general, there is much difference between users, which sensor data they would want to use, see Table 7.5. Again, we asked users to explain their choices. Some people based their decision to collect sensor data on the sensitivity of the particular sensor. Some people based their decisions on a trade-off between usefulness of the data for their personal goal, and how sensitive they find the data. Other users stated that the most important factor for their decision is whether privacy is handled adequately.

Sensor	Real-time measurements	Sampling measurements	Analyze measurements immediately and use only extracted/aggregated data	not
motion sensors	8 (b,c,d,e,f,g,h,i)	0	1 (a)	0
location detection	3 (d,e,i)	3 (c,g,h)	2 (a,f)	1 (b)
light sensor	3 (e,f,h)	3 (c,d,g)	2 (b,i)	1 (a)
sound sensor	2 (e,f)	2 (c,h)	3 (b,d,i)	2 (a,g)
digital communication	2 (e,f)	1 (h)	3 (b,c,d)	3 (a,g,i)
computer interactions	3 (e,f,h)	1 (d)	3 (b,c,i)	2 (a,g)
computer content	1 (e)	1 (h)	5 (b,c,d,f,i)	2 (a,g)
webcam	2 (e,f)	2 (d,h)	2 (b,c)	3 (a,g,i)
Kinect 3D camera	3 (d,f,g)	2 (e,h)	3 (b,c,i)	1 (a)
heart rate sensor	5 (b,d,e,f,g)	2 (c,h)	1 (i)	1 (a)
skin conductance sensor	4 (d,e,f,g)	3 (b,c,h)	1 (i)	1 (a)

Table 7.5: Which sensors may be used to collect data and in what detail. (Amount of users that chose the option denoted, together with user IDs.)

In general, we find that many sensors are perceived as very privacy sensitive (see Table 7.6). Only the motion sensor is experienced as less privacy sensitive, which is in line with the fact that participants judge information on steps/ time active as not very sensitive. The body sensors for collecting heart rate and skin conductance are generally perceived as neutral or only little concerning. In general, many participants express a strong concern for the webcam, sound sensor, computer content and digital communication. Still there is variation between users regarding the exact privacy sensitivity of each sensor. We asked users again to explain what they are afraid of. Users are mainly concerned of misuse of the data, privacy and being controlled and judged.

Sensor	no concern	little concern	neutral	light concern	strong concern
motion sensors	1 (a)	5 (c,b,d,f,i)	2 (e,h)	0	0
location detection	0	1 (i)	0	3 (c,d,h)	4 (a,b,e,f)
light sensor	0	5 (c,d,e,f,h)	1 (i)	1 (b)	1 (a)
sound sensor	0	2 (c,f)	0	2 (b,d)	5 (a,e,g,h,i)
digital communication	0	1 (f)	0	2 (c,d)	6 (a,b,e,g,h,i)
computer interactions	0	2 (f,d)	0	2 (c,h)	5 (a,b,e,g,i)
computer content	0	0	1 (d)	2 (c,i)	6 (a,b,e,f,g,h)
webcam	0	0	1 (f)	2 (c,d)	6 (a,b,e,g,h,i)
Kinect 3D camera	0	2 (c,f)	2 (d,i)	1 (h)	3 (a,b,e)
heart rate sensor	0	4 (b,c,d,f)	2 (e,h)	1 (i)	1 (a)
skin conductance sensor	0	6 (b,c,d,e,f,h)	1 (i)	0	1 (a)

Table 7.6: How privacy sensitive the users consider each sensor. (Note: due to a technical issue responses for participants ‘a’ and ‘g’ on ‘a little sensitive’ and ‘not very sensitive’ were not properly recorded.)

We also asked participants to explicitly choose their preferences for a list of conflicting requirements (identified in Section 7.4). In summary, we can say that different users make different choices in the trade-offs presented to them. For desired functionality the users are willing to hand in a bit privacy or do extra effort. There are no options that all users prefer in general, but we see some trends (see Table 7.7). Most users prefer full control over the system over a system that is as simple as possible. Most users like the idea of combining different types of data for more insight. Most users prefer to store data on a server (such that it is accessible from different locations). For the other settings, choices are more mixed. We also asked the users to explain their choices, i.e. which choices were easy to make, which difficult and why. What makes decisions difficult is the fact that time and knowledge are necessary to take good decisions. A participant suggested to take over settings from ‘power users’ who have delved into various trade-offs. Moreover, some users noted that the system may also change over time, e.g. from a learning period in the beginning to more automation, or from detailed support to providing overviews. We can conclude that not only the setting should be flexible enough to account for different user’s preferences, but the system also should account for changes in preferences over time.

### 7.5.3 Conclusions and discussion

What this user study showed us is, that users are very different with respect to the choices they make regarding data collection for a user model. With respect to subjective perception of information sensitivity, we found that some sensors are in general perceived as more privacy sensitive (e.g. webcam, sound sensor, computer content, digital communication), others as less privacy sensitive (e.g.

OptionA	#users (IDs)	Vs. OptionB	#users (IDs)
Full control over the system and all settings.	7 (a,b,c,e,f,h,i)	A system as simple as possible (settings only require additional effort).	2 (d,g)
Different types of information are combined for more insight (e.g. sleep and manner of working).	8 (a,b,c,d,e,f,h,i)	Different types of information are kept separate (because of privacy).	1 (g)
The information in the system is exact (costs battery or speed of the system).	6 (a,c,f,g,h,i)	The information in the system gives a rough estimate.	3 (b,d,e)
Collect information by means of pop-ups.	3 (c,d,g)	Do not disturb (but system works worse).	6 (a,b,e,f,h,i)
Store data on a server (data is then accessible from different locations).	7 (a,b,c,d,f,g,h)	Store data only locally.	2 (e,i)
Process data on a server.	6 (a,c,d,f,g,h)	Process data locally (costs battery or speed of the system).	3 (b,e,i)
The system provides automatic support.	3 (a,b,h,i)	Choosing when the system provides support.	5 (c,d,e,f,g)

Table 7.7: Users expressed their preference among several trade-offs (left item vs. right item). (Amount of users that chose the option denoted, together with user ids).

motion sensors, heart rate, skin conductance). However, preferences regarding data collection are diverse and depend on the goal for which they want to use the system and the trade-offs they make for themselves regarding privacy. Someone interested in physical activity will be more willing to collect detailed data on movement, whereas someone interested in stress will be interested in collecting work related data.

Our results on user choices and control are also comparable to what Knijnenburg and Kobsa (2013b) (see online Appendix) found in interviews on information disclosure in context-aware recommender systems. They describe that users consider privacy when deciding what requests to answer. The benefits of disclosure are an important factor in deciding to disclose information (as was also found in our study). Users liked to have the option to choose what to disclose. Many users liked to be able to change their disclosure after seeing actual system performance. This may be something also holding true for our system. Moreover, they report that some users dealt with privacy concerns by providing less detailed information, as was found here. Finally, they describe that users' decision depends on the company that provides the system. This is an aspect we did not investigate.

Managing information privacy seems a balancing act between people's internal conflicting requirements. The perception of privacy is based on a cost-benefit analysis, i.e. a privacy calculus (Laufer and Wolfe, 1977). The individual differences in trade-offs may be due to the fact that users have different values. By

addressing values in the requirements engineering process systematically (Harbers, Detweiler, and Neerincx, 2015), sound personalization requirements may be derived and added to the current requirements baseline.

Knijnenburg and Kobsa (2013a) state that “in a privacy-minded design [...], users are free to decide for themselves what data they want to disclose about themselves.”. However, such decisions are complex and burdensome. In their work they investigate the cognitive processes involved in the users’ “privacy calculus”. They present a model showing how personal characteristics lead to a certain perception of a system (e.g. perceived privacy threats), which lead to a certain user experience, which in turn results in users disclosing more or less demographic information and/or context information. Knijnenburg and Kobsa (2013a) also present work on helping users with information disclosure decisions. They investigated whether different types of justifications for data collection (e.g. recommendations will be better, X% of the users allowed us to use the data) would increase the users’ willingness to share demographic and contextual information. They found that the justification for data collection should be adapted to the specific user type (high/low disclosure tendency, male/female) and goal (high demographic disclosure, high context disclosure and/or high satisfaction with the system). In Rainie (2005) respondents expressed their need for easy-to-use technological tools that would provide users with a sense of control, or at least insight in how their information is treated. Several other empirical studies have confirmed that a lower perceived control of personal information release is associated with more privacy concerns (Xu (2007), Hoadley *et al.* (2010), Dinev *et al.* (2013)).

As a final note, actual behavior might deviate from stated attitudes and intentions (Smith, Dinev, and Xu, 2011). We therefore think that letting users try out an e-coach and tracking app for several days gave them an impression on the possibilities of e-coaching and a feeling for being tracked. Based on these experiences we think the participants were able to give a considerate answer on our questions regarding their data collection choices. We presented the choices as settings, rather than opinion questions, in order to push users to actually think of how they configure the system they use.

A limitation of this study is the small number of participants. As participants voluntarily registered to using the coaching and tracking app, the sample may be biased. Moreover, participants were highly educated and had a technical background, which may have led to these particular results. Nevertheless, even in this small group we see a great diversity of preferences, which may be representative in general.

### 7.5.4 Refinement of the user model

Regarding the design of the ‘CAS for stress reduction’, we can now refine the user model, such that it takes into account the privacy wishes of the users (see Figure 7.3). The most important reason for users to collect data (in detail) is for their personal goal or benefit. Only information necessary for the particular goal of the user should be collected. The user should be able to specify his or her goal in the system. The system should then provide information on which information would have to be collected and which sensors should preferably be switched on. The user is in control of which sensor he actually switches on or off. In case he or she does not want to use a particular sensor, alternatives should be provided, e.g. making a more rough estimate based on other data, or providing the opportunity for manual input. The user may control whether he wants real-time data collection or just sampling of data. The algorithms should account for this. Most robust, we think, is to design the algorithms in a way that data over several minutes is aggregated and analyzed. In the real-time setting the user model contains detailed information on the development over the day. In case the user chooses to only sample data, the user model contains a coarse development over the day. This is also in line with the privacy design strategies ‘minimize’ and ‘aggregate’. The system should minimize the data that is collected and aggregate the data to a level that is suitable for the goal. Moreover, the importance of the privacy strategy ‘control’ is highlighted. The user should be able to control which data is collected. Users will probably not be willing to collect just all data, when they have only interest in specific goals.

Regarding the trade-off analysis we see that users differ in the choices they make for themselves. Ideally, the designer of the system should thus not make choices, but let the user select the option that is most appropriate for him or her. In case a user decides that the sensor data should be analyzed and aggregated right away, this may have to be done on the users’ device, otherwise using a server may be more efficient. The user model may be stored locally on the user’s device (more privacy friendly) or on a server (better for functionality). Depending on the amount of information available in the user model, the system can provide specific functionality and personalization. This requires flexibility in the personalization algorithms. Probably it is good to define certain default functionality, for example a general method to provide coaching without any context information (e.g. 3 tips a day at fixed time slots). The timing can then be improved in case additional information is available, e.g. based on the users location, stress level or current task. The more information is available, the better the personalization may work. Probably in the beginning users will want to experiment how much functionality they can gain with disclosing certain



types of data.

## 7.6 Conclusions and Discussion

In this chapter, we presented a case study on the user-centered design of a context aware support system. We now present the main conclusions and discussion on the human-centered development method, the current design rationale, and our user studies.

### 7.6.1 Human-centered development method

The contribution of this work is providing a human-centered development method for context aware support systems. In our work, we refined the situated Cognitive Engineering methodology (Neerincx and Lindenberg, 2008) to combine knowledge on human factors and enabling technology, with user and operational support demands, to specify a design rationale. We applied several complementary design methodologies to define core functions and functional components, as well as a set of requirements for the CAS, which were then further refined in an evaluation study. In our approach, we particularly focused on analyzing user concerns, complementing the analysis with a privacy impact assessment, and suggesting ways to address privacy in CAS. Combining these methods helps to address privacy concerns as an integral part in the design of context aware systems. We recommend this approach for the design of other CAS, in particular:

1. *Addressing human factors knowledge.* By using relevant domain knowledge, theory-based core functions for the CAS can be defined. A stress-model was constructed (mainly based on stress theory of Le Fevre, Kolt, and Matheny (2006)), from which the following 3 core functions were derived: 1) provide insights in the general stress process; 2) help the employee to change working behavior or recover from stress; and 3) change the work situation by means of work support (see Figure 7.2). The theory indicated that also aspects outside of work may be worth considering.
2. *Assessing enabling technology.* Context-driven requirements analysis (Choi, 2007) helps to a) start with thinking about desired services, b) from there investigating the types of context necessary, and c) finally deciding on the specific inference algorithms and sensors to be used. In this way, CAS design is driven by desired functionality. This results in abstract functional components for the CAS. We defined a high-level architecture (Figure 7.5)

with a user model (Figure 7.3) and context inference mechanisms (Figure 7.4). This provides a technical basis for further research and development.

3. *Acquiring user and operational support demands* (Section 7.3).

- *User needs.* A workshop with potential users is a helpful means to find desired functionality. This resulted in 15 functional requirements (see Table 7.1).
- *User concerns.* When a first full system design is specified, users should specifically be asked about hurdles to use the system, e.g. by means of participatory Scenario-Based Design (J. M. Carroll, 2000). This yielded additional (non-functional) system, usability, quality, privacy and security requirements (see Table 7.2).
- *Privacy.* A Privacy Impact Assessment (Wright, 2012) helps to specifically formulate privacy requirements. We identified several privacy requirements that were not mentioned by users themselves. 8 Privacy Design Strategies (Hoepman, 2014) should be applied, to address the privacy requirements. We integrated these into the high-level architecture (see Figure 7.7).

4. *Harmonizing functional and non-functional requirements.* A Trade-off Analysis (see Sutcliffe (2013)) helps to identify requirements that are contradictory (see Table 7.8). In further system design such trade-offs may need to be settled, or, when possible, the user may be left the choice to make this trade-off for him- or herself.

5. *Evaluating for further refinement.* A formative evaluation with an implemented prototype helps to investigate which sensors and context inference users find acceptable, and what kind of trade-offs they make for themselves regarding data collection and privacy. We found individual differences (presented in Table 7.7). Based on these insights, additional requirements can be formulated and the user model can be refined (see Figure 7.3).

Particularly, the combination of these methodologies yielded several important design decisions for our use case. First of all, the defined services for the ‘CAS for stress reduction’ were as well desired by users, as grounded in theory. The user input in particular gave interesting ideas for the CAS, whereas the theory proved helpful to pin-point in which way particular services address the stress problem. Second of all, context-driven requirements analysis helped us to distinguish between context information for the user model, relevant context

attributes and sensors. In the evaluation study, we particularly asked users on their opinion on several types of information and several sensors. As users seem to be interested in several types of information (e.g. work tasks, topics, mental workload), but judge certain types of sensor information as privacy sensitive (e.g. computer interactions, computer content, webcam), we should investigate the use of alternative sensors or process data locally. Third of all, the user study on user concerns confirmed that privacy is an important issue in a 'CAS for stress reduction'. In the evaluation study with a prototype system, we were able to further pin-point which information and sensors are perceived as most sensitive. Finally, the trade-off analysis helped us to identify which (privacy and functional) requirements are conflicting. The evaluation study revealed, that there are no preferences for these trade-offs in general. Ideally, the system should be flexible enough to account for different user's (privacy) preferences.

We have to note that our CAS for stress reduction is still in development and several presented user studies were rather small scale. Ideally a prototype entailing more functionality should be evaluated with a larger and more diverse user group. User's actual usage behavior should be studied, to evaluate the privacy-friendly CAS design.

### **7.6.2 User studies**

Finally, our user studies yielded general insights on concerns and data collection choices in CAS. With respect to user concerns, we found that many concerns are related to privacy, i.e. who can see the data, what will happen with the data, sharing data with other parties, and the feeling of being watched. Moreover, concerns regarding the effort the system would require, highlight the importance of automatic inferences and smart support, while users also wish to stay in control. Combining automatic processing with human interaction seems a good solution. Furthermore, inferences should be reliable, without any negative effects on the interactive response time of the PC. Solutions can be using CPU and memory efficient algorithms, running inference algorithms on a server or analyzing samples of data.

With respect to subjective perception of information sensitivity, we found that some sensors are in general perceived as more privacy sensitive (e.g. webcam, sound sensor, computer content, digital communication), others as less privacy sensitive (e.g. motion sensors, heart rate, skin conductance). However, preferences regarding data collection are diverse and depend on the goal for which users want to use the system and the trade-offs they make for themselves regarding privacy. With respect to privacy-friendly user modeling, the user should be able to specify his goal in the system, and the system should

then provide information on which information would have to be collected and which sensors should preferably be switched on. Users may want to experiment how much functionality they can gain with disclosing certain types of data.

As a limitation, we have to note that several user studies were performed with users with a technical background. This may have led to different demands and concerns as compared to asking users from other backgrounds. Furthermore, the system is developed and evaluated with users in the Netherlands, a country with decent labor laws, which had an influence on the outcomes of our study. In other settings, employees might be less willing to use a context aware system for stress reduction.



## Privacy and user trust in context-aware systems

*Context-aware systems (CAS) that collect personal information are a general trend. This leads to several privacy considerations, which we focus on in this chapter. We present as use-case the SWELL system and address privacy from two perspectives: 1) the development point of view, in which we describe how to apply ‘privacy by design’, and 2) a user study, in which we found that providing detailed information on data collection and privacy by design had a positive effect on trust in our CAS. We also found that the attitude towards using our CAS was related to personal motivation, and not related to perceived privacy and trust in our system. This may stress the importance of implementing privacy by design to protect the privacy of the user.*

---

This chapter is based on Koldijk, Koot, *et al.* (2014). “Privacy and User Trust in Context-Aware Systems”. Published in: Proceedings of User Modeling, Adaptation and Personalization (UMAP, 2014).

## 8.1 Introduction

In this research we want to investigate how to address privacy in CAS and whether information on privacy has a positive impact on users' trust and attitude towards using the system. In Shin (2010) they found that in social networks, privacy and security had an effect on the user's trust in a system and the attitude towards the system, which in turn influenced the intention to use the system. An overview paper (Smith, Dinev, and Xu, 2011) outlines that firms can build trust by implementing fair information practices, communicating a privacy policy explicitly and/or using privacy notices and seals of approval.

We first analyze which privacy aspects are of particular interest in CAS by doing a Privacy Impact Assessment. We make use of a use-case called SWELL, in which work related behavior data is collected with sensors, to provide personalized feedback and support for well-being at work. As the collected data may include rather personal information (e.g. content worked on or facial expressions), interesting privacy aspects arise. This domain distinguishes our research from related research in which privacy is often investigated in context of social networks, user profiling, e-commerce, marketing or mobile location enhanced technologies (Smith, Dinev, and Xu, 2011). We then outline how Privacy by Design (Cavoukian, 2012) can be applied in CAS, resulting in some simple guidelines for developing privacy-friendly CAS. There are many papers on principles for privacy by design, but empirical studies are sparse. Therefore we performed a user study to investigate the effects of privacy by design on users. Our method is similar to the one used in a study on privacy concerns in location-based mobile services (Barkhuus and Dey, 2003): users were presented our envisioned system and were asked to give ratings. Our hypothesis is that when users have access to detailed information on data collection and privacy by design, the transparency of the system is higher and users have less privacy concerns and more trust in the system. As a consequence, we hypothesize, they have a more positive attitude towards using the CAS.

In the remainder of this chapter we first introduce our use-case (Section 8.2). Then we present important privacy aspects (Section 8.3). In Section 8.4, we describe how privacy by design can be applied. We then present results of our user study (Section 8.5). We end with a Discussion (Section 8.6) and Conclusion (Section 8.7).

## 8.2 Context aware system use-case: SWELL

In this section we present a use-case from the project SWELL<sup>1</sup> to apply our analyses regarding privacy to. The SWELL system makes use of a variety of contextual sensors, which makes it interesting for analyzing associated privacy issues. We first outline the CAS and then present a scenario.

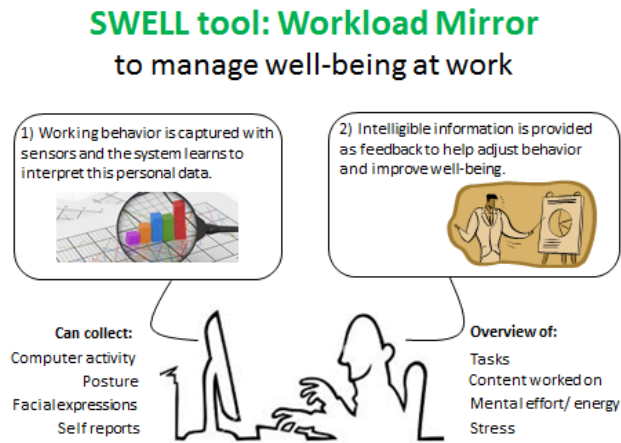


Figure 8.1: Information about the SWELL system.

**SWELL Workload Mirror.** The SWELL Workload Mirror is a CAS under current development that provides information about working behavior to help employees reach more well-being at work (Koldijk, 2012). Knowledge workers often experience stress building up, which in the worst case results in burn-out. We think that helping knowledge workers to become more aware of what makes them feel stressed, can help them handle and avoid stress. The SWELL system senses data about an user's environment with unobtrusive sensors, combined with occasional self-reporting by the user. Smart reasoning algorithms extract the recent context and mental state from this data. The system is aimed at helping users to reach their well-being goals by providing information, feedback and support.

**SWELL scenario.** Bob is 40 years old and works in an office from 9 to 5, where he performs knowledge work. Since some time now, Bob feels some tension and finds it hard to get work off his mind in the evenings. At the end of

<sup>1</sup><http://www.swell-project.net>



his working day he often notices that he has not completed all planned tasks and he feels stressed. Bob decides to use the SWELL system (see Figure 8.1). At the end of his working day he opens the SWELL Workload Mirror to look back at his day. He sees an overview of the tasks he performed and content he worked on, combined with information on his subjective energy level. He notices that he worked very fragmented which probably caused his loss of overview and decline in energy. Bob decides that it would be better for him to stay focused on his planned work and determine a timeslot to do all ad-hoc tasks. He enables a functionality of the SWELL tool, which warns him when he makes too many task switches again. Bob also notices that, in fact, he has done a lot of useful things today and can go home satisfied.

### 8.3 Privacy aspects

To analyse the potential privacy risks around collecting personal data with the SWELL system, we performed a Privacy Impact Assessment (PIA)<sup>2</sup>. As Wright (2012) describes it: “PIAs provide a way to detect potential privacy problems, take precautions and build tailored safeguards before, not after, the organisation makes heavy investments in the development of a new technology, service or product.” (p. 54). We went through the PIA question catalogue and in this section we present the resulting main privacy considerations and provide the most important PIA suggestions to build a privacy-friendly CAS.

- Goal of data collection: We found that it is very important to clearly describe the goal for which the data is collected. Only when users understand what the system does and why the collection of data is necessary, they will be able to take a well informed decision on how to use the system.
- Type of data: The PIA highlighted that the type of data should be suitable to fulfil the goal. Do not collect more data than necessary. Be aware that the combination of different sorts of data can be even more privacy sensitive. Store data as aggregated as possible, for example only store summaries of facial expressions instead of video. Time limit the storage of personal data. This prevents function creep, i.e. using the data for other purposes. In any case, identifiers such as full names and email-addresses should be avoided where possible.
- Reactions to the system: In the PIA it was pointed out that you should be aware that reactions to new innovative systems are hard to predict. The

---

<sup>2</sup>Also reported in Chapter 7.

data that you want to collect can be sensitive, for example when you collect data on geo-location or work performance. Prevent reputational damage. The right story and suitable introduction will be essential to make the tool a success. There is a risk that people involved do not want to participate. For users who do not want to use e.g. a camera an alternative means to get the necessary information should be provided, e.g. let users input their mood themselves.

- **User control:** In the PIA it was recommended to let the user be in control of the system and the settings. You should tell the user which data is collected and (if applicable) who will have access to this data. Their permission should be given based on a free and well informed decision. Giving information on what is done with the data also contributes to transparency and evokes trust. Users have the right to see their own data and may request removal of data.
- **Quality of the data:** The PIA highlighted that it is important to pay attention to the quality of the data. The data should be up-to-date, correct and complete. Depending on the sensor, the data can be more accurate (e.g. computer logging) or less accurate (e.g. facial expressions from video analysis). You can reach better quality by for example letting the user check, correct or update the data. Be aware of consequences of using wrong data.
- **Security of the data:** Security of the data is a must. In the PIA it was recommended to set up a data security plan to establish which security actions are taken to guarantee suitable protection of the data. Prevent unwanted or unauthorized access of the data. Take the sensitivity of your specific data into account.
- **Data responsibilities:** The PIA pointed out that the more parties are involved, the higher the risk of data getting lost, unclear responsibilities or use of data for other purposes. Take care that all parties handle the data carefully. Make a clear data description and a clear description of tasks and responsibilities. Make clear who has to take the measures necessary to prevent risks.
- **Data sharing:** We found that in case of data sharing, you should take care that the user gives consent and that the data is used in the intended way. Data can be shared in several ways. First of all data may be shared between users, when it is the wish of the specific user to share data for a personal goal or benefit. Second of all, data may be shared for improving the system, e.g. to train underlying models with all users' data. Thorough

analysis should be done whether no personal information could leak in this way. Finally, it may be interesting to share collected datasets with the research community. When data is distributed you should describe the data well and take care that the distribution of data is in line with the expectations of the users involved. Make a clear data description document. Pay attention to purpose limitation and risks resulting from combining data from different sources.

Being aware of these points of interest at an early stage of design should enable developers to implement privacy into their context aware system.

## 8.4 Privacy by Design

In this section we present how the outlined privacy aspects for the SWELL use-case can be addressed from the developers perspective by using Privacy by Design (Cavoukian, 2012). We describe how 8 Privacy Design Strategies (Hoepman, 2014) can be applied to develop a context aware system that follows current privacy legislation. We also give some tips on specific Privacy Design Patterns that can be used to implement each strategy. For a more elaborate description and specific references refer to Bodea *et al.* (2013).

- First of all it is important to INFORM the user about the goal of the system and the data that will be collected for this aim. You should always use Informed Consent, which means that the you get permission from the user to collect data for a specified purpose. You can also provide the user a Privacy Dashboard, such that the user has an overview over his privacy settings.
- Moreover, it is important to give the user CONTROL over the data and what is done with it. There are different ways to let users feel in control. Information helps users to understand the system and power allows them to decide which data is collected, how it is used and with whom shared. Offering Privacy Choices helps to give them a feeling of control and a system that is easy to use also increases the perceived control.
- The task of the designer of the system is it to MINIMIZE the amount of data that the system stores. This can be accomplished by selecting only the most relevant features (e.g. storing facial expression features instead of raw video recordings). In any case it is a good idea to only use Pseudonyms as identifiers, instead of storing data together with the users' real names. Furthermore, take care of good Anonymization. Even when

you do not store the user's name, the unique combination of e.g. age and GPS location can make a user of the system identifiable. Prevent having identifiable entries and use k-anonymity. This means making at least k entries identical, for example by aggregating "age = 22" to "age = 20 to 30".

- By applying reasoning the data can often be AGGREGATED even further. Instead of detailed features, inferred information can be stored (e.g. whether someone experienced stress or not, instead of all facial expressions). You can for example Aggregate Data over Time, e.g. the main application of the last 5 minutes or the main facial expression. This also lessens the amount of data the system has to handle. Moreover you can also Blur Personal Data. This means you provide personal data only in a detail that is necessary and blur the rest, e.g. store location information not as a coordinates, but as a city name.
- The developer should take care to HIDE personal information, such that the data strictly belongs to the user and cannot be seen by others. When a user or application wants to access the data, Authentication should be used to ensure that no unauthorized access to the data takes place. To ensure the security of the data it is a good idea to Store Data Encrypted. You should encrypt the data locally on the users device and then send it over a secure connection to the cloud for storage. When the aim is to publish (parts of) the data one could apply Sampling. Instead of releasing all data a sample is drawn for releasing on the (public) cloud.
- Moreover it might be useful to SEPARATE different sorts of data. Storing data from different individuals at separate locations is called Horizontal Data Separation, while storing features in separated locations is called Vertical Data Separation. When handling privacy sensitive data it is also good to apply Decentralization and store (parts of the) data only locally, on the user's device.
- The system should be able to ENFORCE and DEMONSTRATE that it fulfils current legislation around privacy. You might want to use Sticky Policies, especially when sharing data. This means that you store alongside with your data its privacy policy for handling this data. In this way you prevent wrong use by 3rd parties.

By applying these 8 Privacy Design Strategies in the development of a CAS the resulting system will be privacy-friendly by design, adhering to current legislation.

## 8.5 User perspective: Evaluation study

Now we have seen how privacy can be addressed in the development of a CAS, we want to evaluate what effect giving information on privacy by design has on users. Our hypothesis is that when users are better informed about the data collection and privacy by design, the transparency of the system is higher and users have less privacy concerns and more trust in the system. As a consequence they have a more positive attitude towards using the system (see Figure 8.2 for our expected model). We also think personal characteristics play an important role. General privacy concerns might have an influence on perceived privacy and trust in a new system. Personal motivation might have an influence on attitudes towards use of the system. In the remainder of this section we outline how we tested our hypotheses in a user study with a mock-up of our SWELL tool.

### 8.5.1 Method

**Participants** 124 people participated in our user study, 60% male, with an average age of 38 ( $SD = 10.6$ ). Colleagues from other TNO departments (technical and behavioural sciences) were invited as participants, as they are knowledge workers and potential users of the SWELL tool to improve well-being at work. On a scale from 1-7 our participants scored on average slightly positive on well-being (4.7,  $SD = 1.2$ ) and slightly positive on the item 'I want help to improve well-being' (Motivation) (4.9,  $SD = 1.7$ ). Moreover, they scored on average neutral on Privacy concerns (4.1,  $SD = 1.5$ ).

**Design** We manipulated whether the participants did or did not get extra information on data collection and privacy by design. Our experiment had thus a between-subject design and our independent variable is 'privacy information' (no, yes).

**Procedure** An email was sent out to various TNO departments. By clicking a link, the participant was randomly assigned to the condition with or without privacy information and shown a website. On this website, first a short presentation was shown, either with or without slides on privacy. Both groups were then asked to fill in the same questionnaire.

**Materials Presentation.** The first 7 slides were the same for both groups and presented a scenario for the SWELL Workload Mirror (see Section 8.2). Both groups were told that the goal of the SWELL tool is to support self-management

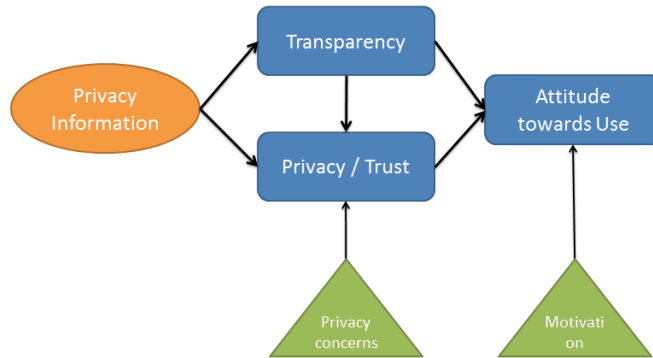


Figure 8.2: Expected model. We manipulated whether participants had access to extra Privacy Information. Our 3 dependent variables are Transparency of the SWELL tool, attitudes regarding Privacy and Trust, and Intention to Use the SWELL tool. We expect that also personal characteristics (Privacy concerns and Motivation) play a role.

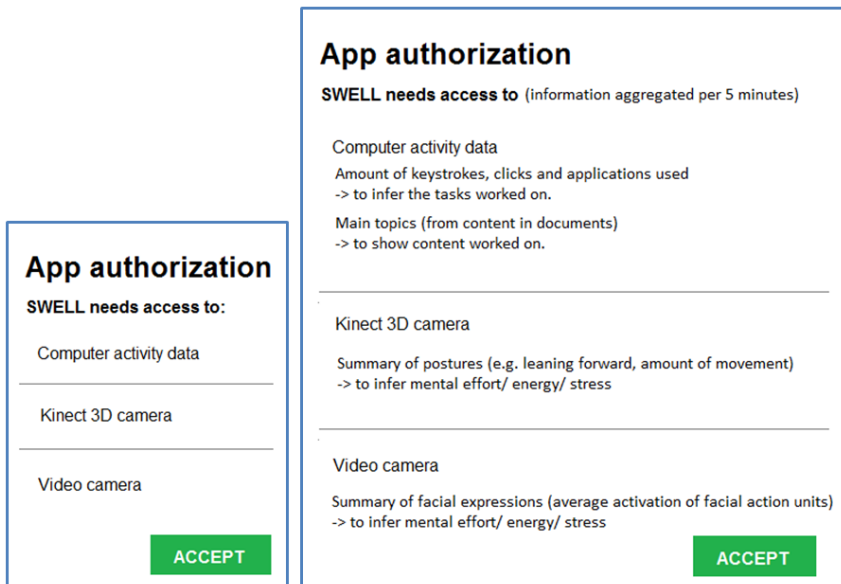


Figure 8.3: Mock-up of the access rights dialogue. Left: Control condition, right: Privacy condition.

of stress and that the users could enable or disable functionalities as they wish, such that the SWELL tool optimally supports them with functionality that they desire (e.g. sharing information with others).

The privacy group got to see extended information on the data that the system would collect (see Figure 8.3). Moreover, an additional slide gave them the following information on privacy by design:

- Purpose limitation: The collected data is only used for giving yourself insights to enable self-management.
- Control: You can enable or disable the computer logging, camera or Kinect sensors.
- Data minimization: The tool only processes data that is necessary to provide the functionality that you desire, e.g. the tool will use document content only when you want an overview of topics worked on.
- Data aggregation: The sensor data is processed locally on your device. Only summary information, such as topics, average posture or facial expression, is stored - no keystrokes or video.
- Adequate protection: Your data is hidden from unauthorized access.
- Data subjects right: You have full control over your data, can view or delete it.

**Questionnaire.** The questionnaire had items on the following main categories: transparency of the SWELL tool, perceived privacy and trust, and attitudes towards use of the SWELL tool (see Figure 8.4, items partly adapted from Shin (2010)). Besides these main items of interest, we added some items on personal characteristics. We used 7-point Likert scales (1 = 'not' to 7 = 'very much').

**Dependent variables** To determine the main underlying concepts of the questionnaire items, we performed a factor analysis (PCA, see Figure 8.4). We found 3 main underlying components, which represent: 'Transparency', 'Privacy/ Trust' and 'Attitude towards Use'. To test the reliability of each scale, we calculated Cronbachs' alpha (coefficient of internal consistency) for each set of items. As for all 3 scales alpha was high enough ( $> .6$ ), we computed sum scores of the sets of items and averaged them to yield 3 main dependent variables.

Questionnaire item	C 1	C 2	C 3	Concept	$\alpha$
I would trust SWELL to protect my privacy. *	.862	.176	.173	<i>Privacy/ Trust</i>	.884
I am confident that the information I provide will be secure. *	.861	.144	.105		
(reversed) I am concerned that the information I collect could be misused. *	.832	-.049	-.055		
(reversed) I believe inappropriate parties may view the information I provide. *	.795	-.022	-.048		
The SWELL tool is a trustworthy system.*	.696	.202	.236		
I feel in control of my personal data.	.645	-.027	.481	<i>Attitude towards Use</i>	.890
The thought of using the SWELL tool is appealing to me. *	.071	.923	.053		
I think using a computer system like SWELL could help me improve my well-being at work.	.013	.834	-.020		
I have positive feelings towards the SWELL tool in general.*	.231	.802	.127		
I think benefits of using the SWELL tool outweigh potential costs (effort, time).	-.052	.795	.027		
I want to use the SWELL tool.	.104	.783	.038	<i>Trans- parency</i>	.655
I understand which information will be collected using the SWELL tool.	.146	.022	.836		
It is clear to me how the SWELL tool works.	.050	.102	.830		
My current level of well-being at work (e.g. feeling in control, productive, energetic) is...				Personal characteristics	
I want help to improve my well-being.					
In general, I would say I am concerned about my privacy.					

Figure 8.4: Questionnaire items (\* item adapted from Shin (2010)), loadings on PCA components (with Varimax rotation), and Cronbachs' alpha for combining these items to one concept.

## 8.5.2 Results

**Personal characteristics** As we think personal characteristics may have an important influence on our dependent variables, we calculated Pearson correlations to check for these dependencies. We found a significant moderate correlation between Privacy Concerns in general and perceived Privacy/ Trust ( $r = -.548$ ,  $p < .001$ ). People who in general have many privacy concerns tend to score low on perceived privacy and trust regarding the SWELL tool. Furthermore, we found a significant weak correlation between the level of well-being and the desire to improve well-being ( $r = -.337$ ,  $p < .001$ ), as well as a significant moderate correlation between the desire to improve well-being and Attitude towards Use of the SWELL tool ( $r = .457$ ,  $p < .001$ ). This means that people with low well-being want to improve well-being more, and people who want to improve well-being more have a more positive attitudes towards using the SWELL tool. In the remaining analyses we will use these personal characteristics as covariates.



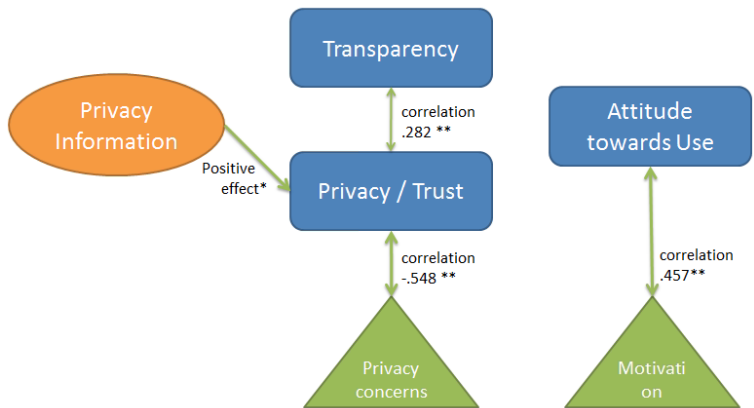


Figure 8.5: Results. Influence of privacy information was analysed with ANOVA's. Relations among variables were analysed with Pearson correlations. (\* significant on the .05 level, \*\* significant at the 0.01 level).

**Effects of privacy information** We were interested in whether giving extra information on data collection and privacy by design would have a positive impact on Transparency of the SWELL tool, attitudes regarding Privacy and Trust, and finally on Attitude towards Use of the SWELL tool (see Figure 8.2 for our expected model). Therefore, we performed an ANOVA with privacy information (yes, no) as between-subject factor and Privacy/Trust as dependent variable, using the personal characteristic Privacy Concerns as covariate. We found a significant effect of privacy information on Privacy/Trust ( $p = .049$ ). As expected, privacy information had a positive effect on attitudes regarding privacy and trust in the SWELL tool ( $\text{avg}(\text{control}) = 3.85$  vs.  $\text{avg}(\text{privacy}) = 4.24$ , see Figure 8.5). Moreover, we performed an ANOVA with privacy information (yes, no) as between-subject factor and Attitude towards Use as dependent variable, using the personal characteristic Motivation as covariate. We did not find a significant effect of privacy information on Attitude towards Use ( $p = .616$ ,  $\text{avg}(\text{control}) = 4.03$  vs.  $\text{avg}(\text{privacy}) = 4.16$ ). We also did not find a significant effect of privacy information on Transparency ( $p = .332$ ,  $\text{avg}(\text{control}) = 4.64$  vs.  $\text{avg}(\text{privacy}) = 4.86$ ).

To further investigate the relationships between our 3 dependent variables we calculated Pearson correlations. We found a significant weak correlation between Transparency and Privacy/ Trust ( $r = .282$ ,  $p = .001$ ), meaning that a high score in transparency is slightly related to a high score in privacy and trust. We did neither find a meaningful correlation between Transparency and Attitude to-

wards Use ( $r = .132$ ,  $p = .143$ ) nor between Privacy/ Trust and Attitude towards Use ( $r = .170$ ,  $p = .059$ ).

**Summary** From our analyses we can conclude that attitudes regarding Privacy and Trust in the SWELL tool are moderately related to personal Privacy Concerns. Moreover, giving users privacy information seems to have a positive effect on perceived Privacy and Trust regarding the SWELL tool. We found that the Attitude towards Use of the SWELL tool is moderately related to personal Motivation, and contrary to our expectations, not related to attitudes on Privacy and Trust.

## 8.6 Discussion

Our first hypothesis was that when users have access to detailed information on data collection and privacy by design, they have less privacy concerns and more trust in the system. This hypothesis was confirmed in our user study. So, to build trust in your CAS it is a good idea to communicate information about data collection to the user and to address privacy.

Our second hypothesis was that the consequence of more trust would be that users would have a more positive attitude towards using the CAS. This hypothesis was not supported by our data. We found that users base their attitude and intention to use the system mostly on the added value it has for them, and privacy and trust considerations might not be obvious or important enough to be taken into account. This has previously been found and termed the 'privacy paradox': people disclose personal information despite their privacy concerns (Compañó and Lusoli, 2010). We might see the consequences of this when users mindlessly accept all access rights in order to use a desired app. 'Privacy calculus' states that consumers weigh the risks against the benefits of disclosing information (Smith, Dinev, and Xu, 2011). As far as users might be underestimating the risks, Rubinstein (2012) suggests that responsibility for correct data usage should shift towards companies and away from users, who are often left in the dark after consenting to something they may not have read in full detail or understanding. Research has also shown that although people desire full control over their data, they favor technical and other supply-side solutions ('control paradox' (Compañó and Lusoli, 2010)). Therefore we think it is important to implement privacy by design to adequately protect the privacy of the users.

In April 2016 the new General Data Protection Regulation (GDPR, EU (2016)) was adopted, which will enter into application in May 2018. Its primary objective is to give users back control over their data and unifying the regulations

within the EU. The regulation explicitly states that Privacy by Design is required, ensuring that data protection is designed into the system. It also states that Data Protection Impact Assessments need to be conducted when specific risks occur. Our research is thus a step towards these new requirements. Non appliance results in a fine up to 20.000.000 Euro or up to 4% of the annual turnover.

We want to note that due to our methodology (using a presentation to outline the system and a questionnaire to assess the users' attitudes) only first insights can be gained. Ideally, users should be asked to really install the CAS to do a more thorough analysis on the relation between perceived privacy and actual use of the system, which might deviate from stated attitudes and intentions, as pointed out by Smith, Dinev, and Xu (2011).

## 8.7 Conclusions

In this chapter we addressed privacy and user trust in context aware systems (CAS), based on our SWELL use-case. As our SWELL system is a typical CAS in which context data is collected to provide the user with a service, the insights gained are also applicable to other CAS. In the first part of this chapter, we found by means of a Privacy Impact Assessment the following important privacy aspects to address in CAS: Goal of data collection, Type of data, Reactions to the system, User control, Quality of the data, Security of the data, Data responsibilities and Data sharing. We outlined how these issues can be addressed from the developers side by presenting guidelines for Privacy by Design, which can be found in section 8.4.

In the second part of this chapter we presented a user study, in which we found that privacy information had a positive effect on perceived privacy and trust in our system. We also found that the attitude towards using our system was related to personal motivation, and not related to perceived privacy and trust. Therefore we think it is important to implement privacy by design to adequately protect the privacy of the users in context aware systems.

---

## Conclusions and reflection

In this thesis, we investigated how to design a pervasive context aware support system aimed at improving well-being at work. We took a human-centered development approach, in which we applied the situated Cognitive Engineering method to combine theory on work stress with technological possibilities, taking in mind input on user needs. This thesis was therefore divided into three parts: 1) Theory, 2) Technology, and 3) Users. We now present our conclusions on these three aspects. We end this thesis with a reflection.

### 9.1 Theory: Work stress and interventions

The general objective of the system is to improve well-being at work. In the first part of our research, we delved into work stress and intervention theory. We formulated an answer to our first research question: **RQ1** *‘Which concepts are relevant with respect to well-being at work?’* We identified the concepts ‘stress’ (Selye, 1956) and ‘engagement’ (Maslach and Leiter, 2008). An (environmental) stressor can cause a particular perception of the stressor in the individual. This can lead to acute physiological stress responses and, in the long run (due to lack of recovery) to long-term physical, cognitive, emotional and behavioral stress consequences. Engagement (opposite of burn-out) describes a more long-term state, and is measured along three dimensions: energy, involvement and efficacy (or alternatively: absorption).

We also answered our second research question: **RQ2** *‘Which person, work and context conditions can lead to negative stress?’* There are no specific personal, work or context conditions that generally lead to stress. Work becomes stressful when high demands are combined with: insufficient resources; little rewards; little recovery; or an environment that mismatches with personal characteris-

tics. We identified the Job Resources-Demands model (Demerouti, A. B. Bakker, Nachreiner, *et al.*, 2001) and the Effort-Recovery model (Meijman *et al.*, 1998) as most useful models for developing technology-based interventions.

Moreover, we answered our third research question: **RQ3** ‘*How can sensors be applied to automatically infer stress and the context in which it appears?*’ Based on theory, we identified three different aspects that can be quantified: a) Characteristics of the work itself can be measured, e.g. work tasks or topics; b) acute stress can be measured, e.g. heart rate variability; and c) long term effects of stress or recovery can be measured, e.g. sleep or physical activity. In Section 9.2 we present in more detail how sensor data can be used for recognizing tasks, giving insight in work behavior related to stress, and automatically estimating mental states.

Finally, we answered our fourth research question: **RQ4** ‘*Which interventions can be provided by means of pervasive technology to help a knowledge worker improve his well-being at work?*’ In general, three stress prevention approaches are distinguished, aimed at different stages in the stress chain (Ivancevich *et al.*, 1990). Technology can thus either address the stressor (e.g. by providing work support), address short-term stress reactions (e.g. by enhancing coping), or address long-term stress consequences (e.g. helping to improve recovery). Suitable behavioral change techniques (Michie *et al.*, 2008) should be used to address the motivation, ability or trigger to take action (e.g. self-monitoring and reminders to action).

Based upon these insights, we created a comprehensive and practical framework, which relates concepts from various stress and intervention theories, and integrates possibilities for real-time sensor based measurements and interventions with context aware support systems. This framework provides a structure to develop pervasive technology for improving well-being at work grounded in theory.

We used this framework to derive requirements for pervasive well-being technology and presented four prototypes that were implemented: The SWELL Workload Mirror provides an activity and workload overview, designed to find stress sources. The SWELL Fishualization provides department wide feedback for peer support, designed to improve coping. The SWELL NiceWork e-coach provides well-being tips, designed to improve coping or recovery. Last, the SWELL Happy Worker system provides personalized search, designed to support work. Evaluation studies showed that potential end users are positive about the prototypes.

Finally, we provided six key research challenges that were identified in the area of pervasive systems for well-being: a) To develop context aware support systems, multi-disciplinary, theory and data-driven research and development

is necessary. b) Sensor data is relatively easy to collect, the challenge is making sense of this data. c) Research should be done on how longitudinal patterns in sensor data relate to long-term stress consequences and burn-out. d) The most suitable manner for pervasive technology to interact with an employee is a challenging question for human-computer interaction research. e) A system that collects personal data raises privacy concerns, which need to be addressed. f) Measuring and trying to change the behavior of individuals poses ethical questions. This thesis focused on addressing research challenges a) multi-disciplinary research, b) sensor data interpretation and e) privacy.

### **Conclusions and future challenges/ opportunities**

In general, we can say that new technologies bring new possibilities. The rather abstract models can be operationalized to include directly measurable concepts: work characteristics (e.g. task and content worked on), acute stress (e.g. heart rate variability and skin conductance), or long-term stress/ recovery (e.g. sleep time and physical activity). Also new types of support are possible, e.g. providing work support by filtering emails, fostering support by colleagues with a department-wide feedback board, or supporting a work-rest balance with an e-coach. New technology can also be used to directly evaluate the effect that an intervention has on these concepts. Sensors can be used to investigate in how far interventions are indeed followed up (e.g. whether users take a break or become physically active after a suggestion by an e-coach). Moreover, the effects of an intervention can be measured (e.g. whether provided information support indeed decreased mental effort and stress).

In their paper on technology for well-being, IJsselstein *et al.* (2006) state that using technology for improving well-being has many advantages, e.g. its persistence or objectiveness, the possibility to provide just-in-time notifications with relevant, actionable information or their supportive and motivating role. Ideally, the strengths of the technology and the strengths of the human should be combined. In this sense, the SWELL Workload Mirror can provide an objective overview, whereas the user can interpret the data in order to find causes of stress. The SWELL NiceWork e-coach can be persistent in providing well-being tips just-in-time, and the user can take action and feel supported to change his behavior. However, the role of the system and the user should be clear. The user should for example feel in control of the technology, and the system should provide support, while not irritating the user. The most suitable manner to interact with an employee during work is a challenging question for human-computer interaction research.

To conclude, technical experts and social scientists should aim to work to-

gether. It is therefore necessary that the experts understand each others domains well. This thesis provides a starting point to create links between these worlds.

## 9.2 Technology: Inferring working context and mental states

In the second part of our research, we investigated technological possibilities of inferring the user's current working context and mental state from unobtrusive sensors (Chapters 3, 4, 5 and 6).

**Dataset.** First of all, to develop algorithms for inferring working context and mental states, we were in need of a good dataset. We presented our data collection experiment, which overcomes drawbacks that are typically observed in related studies (Chapter 3). Instead of a rather artificial task, 25 people performed typical knowledge work (i.e. writing reports, making presentations, reading e-mail, searching for information), while their working conditions were manipulated with realistic stressors (email interruptions, time pressure). Instead of expensive and/or obtrusive equipment, we used a varied set of (unobtrusive and easily available) sensors: computer logging, video, Kinect 3D and body sensors. We preprocessed the collected sensor data and extracted features per minute: computer interactions, facial expression, body postures, and physiology (heart rate (variability) and skin conductance). The resulting affective and behavioral SWELL-KW dataset is shared with the scientific community in raw and preprocessed form (archived sustainably at DANS: <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:58624>). This dataset has similarities to the publicly available multimodal dataset described by Soleymani *et al.* (2012), which was collected in context of watching emotion inducing video clips and consists of: face videos, audio signals, eye gaze data and physiological signals (EEG, ECG, GSR, respiration amplitude, skin temperature). We demonstrated that analyses of the SWELL-KW dataset yield insights in the effects of stressors at work: Under time pressure, participants experienced significantly higher temporal demand and higher arousal. Email interruptions yielded reports of more mental effort, more positive valence and more dominance. We also found relations between subjective ratings and the sensor data: Explorative correlation analysis showed moderate correlations between mental effort and several facial features. The dataset is therefore a valuable contribution to research fields like work psychology, user modeling and context aware systems.

**Task recognition.** Furthermore, we investigated the following research question: **RQ5** ‘Can knowledge workers’ tasks be recognized based upon computer interactions?’ (Chapter 4). We used real-world office worker data. What we found

is that task recognition must take into account personal factors, as individual users have different work styles and task mixes. Nevertheless, our experiments demonstrated that on an individual basis, tasks can be recognized with an accuracy of up to 80%, which is reasonably high, considering 12 possible task labels. Since different users show different patterns of behavior when performing a task, the classification model should be trained for each specific user to yield optimal task recognition. We found that no more than 2.5 hours (30 instances) of representative training examples is required to train a good model for this task. Unlike other research, in which clearly structured tasks were modeled (see e.g. Natarajan *et al.* (2008)), our research has shown that task recognition also works for less structured tasks and more spontaneous activity.

**Visual analytics.** Moreover, we described how we applied visual analytics to the SWELL-KW work behavior dataset (Chapter 5). We addressed the research question: **RQ6** *‘How can sensor data be used to gain insight into work behavior, specifically related to stress at work?’* We found that mental effort seems to be most closely related to facial expression features. There are, however, many individual differences. By means of a heat map we were able to visualize meaningful patterns in facial activity for an individual user. The visualization was made more insightful by rendering facial expressions on an avatar. Finally, we identified several facial expressions that are typically related to a low or high mental effort. We conclude that facial expressions may be a promising measurable outward characteristic that can be visualized to indicate mental state patterns during work. The benefit of incorporating visual analytics to our problem, instead of a black box machine learning approach, was to gather a deeper understanding of the structures in our data and to gain insights from individual users’ data. To our knowledge this type of analysis has not been applied before on multimodal recordings of naturalistic work behavior data.

**Detecting stress.** Finally, we present work on detecting stress in offices (Chapter 6). In our work, we addressed two methodological and applied machine learning challenges: a) Detecting work stress using several (physically) unobtrusive sensors, and b) Taking into account individual differences. We answered our research question: **RQ7** *‘Can we distinguish stressful from non-stressful working conditions, and can we estimate mental states of office workers by using several unobtrusive sensors?’* A comparison of several classification approaches showed that neutral and stressful working conditions can be distinguished with about 90% accuracy. Posture yields most valuable information, followed by facial expressions and physiology. Furthermore, we found that the subjective variable ‘mental effort’ can better be predicted from sensor data than e.g. ‘perceived stress’. A



comparison of several regression approaches showed that mental effort can best be predicted by a model tree (correlation of 0.82). Facial expressions yield most valuable information, followed by posture.

Instead of only measuring physiological stress reactions (Matthews, McDonald, and Trejo, 2005), focusing on behavioral cues like facial expressions or postures (Pantic *et al.*, 2007), or merely using computer interaction data (Vizer, Zhou, and Sears, 2009), we investigated features from different modalities to find the strongest indicators of the users' mental state.

We also answered our research question: **RQ8** '*How important are individual differences?*' With respect to estimating mental states, we see that information on the participant is important enough to be selected as one of the best features. We further find that a general model tested on a new user does not perform well. This suggests that especially for the task of estimating mental states it makes sense to address individual differences. When we train models on particular subgroups of similar users, (in almost all cases) a specialized model performs equally well or better than a general model.

### Conclusions and future challenges/ opportunities

In general, we can say that sensor data is relatively easy to collect, the challenge is making sense of this data. People differ in their (work) behavior, so there is a need to build personalized models. This brings methodological challenges that need to be addressed, e.g. how to instantiate a model for a new user. Ideally, a system should be able to give a warning, in case it predicts that the current behavioral pattern will cause long-term problems, like burn-out. Therefore, research should be done on patterns in sensor data over time and how they relate to long-term stress building up. Finally, the relation between objective measures based on sensor data, and subjective experience based upon our own feelings should be investigated.

### 9.3 Users: Needs and concerns

In the third part of our research, we investigated human-centered design for effective and user-friendly context aware support systems (Chapters 7 and 8). We addressed the following research question: **RQ9** '*How can we refine the 'situated cognitive engineering' methodology (Neerincx and Lindenberg, 2008) on two aspects: a) defining the context during the requirements engineering process, and b) addressing functional and non-functional requirements coherently?*' We combined several complementary design methodologies. In this way, we were able to define core functions and functional components, as well as a set of requirements for the

context aware system (CAS), which were then further refined in user study. In our approach, we particularly focused on analyzing user concerns, complementing the analysis with a Privacy Impact Assessment (Wright, 2012), and suggesting ways to address privacy in CAS. Combining these methods helps to address privacy concerns as an integral part in the design of context aware systems. In April 2016 the new General Data Protection Regulation (GDPR, EU (2016)) was adopted, which explicitly states that Privacy by Design is required and that Data Protection Impact Assessments need to be conducted when specific risks occur. Our research is thus a step towards these new requirements.

Moreover, we presented work on privacy and user trust in context aware systems (Chapter 8). We first answered the research question: **RQ10** *‘How should privacy be addressed in the design of CAS?’* We performed a Privacy Impact Assessment (Wright, 2012) and outlined how Privacy Design Strategies (Hoepman, 2014) can be applied to address the identified privacy issues in CAS.<sup>1</sup> The strategies ‘minimize’, ‘aggregate’ and ‘control’ are particularly interesting in the view of CAS and pose interesting new opportunities and challenges. We also presented a user study, in which we addressed the research question: **RQ11** *‘What effect does information on privacy by design have on users?’* Our first hypothesis was: a) ‘When users have access to detailed information on data collection and privacy by design, they have less privacy concerns and more trust in the system’. This hypothesis was confirmed in our user study. So, to build trust, it is a good idea to communicate information about data collection to the user and to address privacy. Our second hypothesis was: b) ‘Users have a more positive attitude towards using the CAS, as a consequence of increased trust in the system.’ This hypothesis was not supported by our data. We found that users base their attitude and intention to use the system mostly on the added value it has for them, and privacy and trust considerations might not be obvious or important enough to be taken into account. This has previously been found and termed the ‘privacy paradox’ (Compañó and Lusoli, 2010): people disclose personal information despite their privacy concerns. Therefore, it is important to implement privacy by design to adequately protect the privacy of the users in context aware systems.

### **Future challenge/ opportunities**

In general, we can say that the success of pervasive systems collecting context data depends on the acceptance by users. A system that collects personal data raises many privacy questions. Therefore, privacy should be integral part of the design process (e.g. doing a Privacy Impact Assessment or implementing

---

<sup>1</sup>This work was also presented in summarized form in Chapter 7

Privacy by Design). Currently, legislation around data protection is changing. The consequences for the architecture and functionality of CAS should be investigated. Moreover, measuring and trying to change the behavior of individuals poses all kinds of ethical questions. Is it acceptable to monitor and change the behavior of an employee? It is difficult to predict how such new pervasive e-coaching systems will be perceived and used (or even misused) when applied in real-world work settings.

## 9.4 Reflection

We now present a reflection on the research challenges that we aimed to address, and the limitations of our work. Then, we reflect on the domain challenge of addressing work stress, and provide an outlook for the future.

### 9.4.1 Addressed research challenges and limitations

With this thesis, we contributed to a better understanding of how theoretical (work stress) models can be operationalized, and how sensors can be used to infer context aspects (relevant to work stress). Moreover, we contributed to a better understanding of how theoretical and technological insights can be combined with input on user needs and concerns, to build effective, innovative and user-friendly support systems. More specifically, we addressed the identified challenges in the following three research areas:

1. Behavioral/Social Sciences and Occupational Health: We extended existing theoretical work stress models with new technological operationalizations.
2. Computer Science and Artificial Intelligence: We demonstrated the potential of several sensors to infer relevant aspects of the user's current working context.
3. User-oriented Design and Requirements Engineering: We showed how to harmonize functional requirements and privacy requirements in the design of context aware systems.

Regarding our **theoretical framework**, the biggest challenge was the vast amount of available concepts and models regarding well-being at work. We had to make choices on what concepts and theories to include. Our selection may reflect our specific scoping. We focused on providing a general and simple overview, combining different areas of research.

Regarding **data collection**, we have to note that simulating a realistic work setting and inducing stress was challenging. Real-world stress is complex, and may involve worries or things outside work and stress building up over days. Therefore, a limitation of our collected dataset is that only short term effects of stressors can be investigated. Regarding **task recognition**, clearly not all work of a knowledge worker can be captured on basis of computer activity, e.g. time spent in meetings, phone calls, talks with colleagues or reading printed documents. Regarding **visual analytics**, a user-study should be performed to evaluate whether the resulting visualizations can indeed help employees to detect alarming patterns in work behavior. Regarding **detecting stress**, the data that we used here, was captured in a realistic office setting in an experimental context, which means that the quality of all recordings was high. In a real-world office setting, recordings may be more noisy.

Regarding the **human-centered design method**, we have to note that several user studies were rather small scale. Ideally a prototype entailing more functionality should be evaluated with a larger and more diverse user group. Moreover, the system is developed and evaluated with users in the Netherlands, a country with decent labor laws, which may have had an influence on the outcomes of our study. Regarding our **user study on trust**, we have to note that due to our methodology (using a presentation to outline the system and a questionnaire to assess the users' attitudes) only first insights can be gained. Ideally, users should be asked to really install the CAS to do a more thorough analysis on the relation between perceived privacy and actual use of the system, which might deviate from stated attitudes and intentions.

### 9.4.2 Support technology to manage stress at work

In this thesis we aimed to address work stress by means of new pervasive (sensing) technologies. The part on work stress theories showed us that there are many possible ways in which technology can support employees to address work stress. Our work on technical possibilities showed that interpreting real-world sensor data in terms of mental states is still challenging, e.g. due to individual differences. The part on user needs showed us, that privacy is currently an important concern of users. Taking all these insights together, I personally think that, to address work stress within the next few years, we should aim for rather simple, but effective technical solutions. Regarding the challenges in sensor data interpretation and the feasibility of collecting data from users who are concerned about privacy, I would rather not suggest to aim for sophisticated user models. I expect that combining simple forms of context recognition with an e-coach designed in a smart way and specific user input can certainly help

people, like Bob from the Introduction, to reflect on their behavior and provide support for coping with work stress.

However, it is hard to make predictions for the future. At the start of this project, 4 years ago, I was concerned that many interesting aspects related to stress could not be measured in the real-world or that no-one would want to measure e.g. his body signals during work. Time showed me that trends in technology are entering the consumer market very quickly, with Kinects and smart watches being used. What at the start of the project was done only by some enthusiast quantified-selfers, is now done by my own family: tracking activity by measuring steps, and using this to improve health. Maybe this is a current trend, which will subside when users got those self-insights and start focusing on something else again that is more interesting. Or the hype continues, and people become more and more interested in new sensing devices, gadgets and tracking more and more aspects of their lives. I could imagine many big companies being interested in measuring not only how people interact with their services, but also how people feel or think. It is still questionable in how far consumers are interested enough in new sophisticated services, in change for personal data. I personally expect that most people won't refuse something that makes life easier. The interested reader is referred to the fiction novel 'Free to fall' by L. Miller (2015):

"What if there was an app that told you who to date, what songs to listen to, what coffee to order, what to do with your life - an app that was guaranteed to ensure your complete and utter happiness? What if you never had to fail or make a wrong choice? What if you never had to fall?"

This fictitious app builds sophisticated user models and optimizes an entire populations' lives. Who would not want a perfect life? Well, as you might figure, it is questionable whether this turns out well.

# Bibliography

- Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., and Steggles, P. (1999).** "Towards a better understanding of context and context-awareness". In: *Handheld and ubiquitous computing*. Springer, pp. 304–307 (cited on page 130).
- Acquisti, A. (2004).** "Privacy in electronic commerce and the economics of immediate gratification". In: *Proceedings of the 5<sup>th</sup> ACM conference on Electronic commerce*. ACM, pp. 21–29 (cited on page 132).
- Acquisti, A. and Gross, R. (2006).** "Imagined communities: Awareness, information sharing, and privacy on the Facebook". In: *Privacy enhancing technologies*. Springer, pp. 36–58 (cited on page 132).
- Adams, A. (2000).** "Multimedia information changes the whole privacy ballgame". In: *Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions*. ACM, pp. 25–32 (cited on page 153).
- Albrecht, D. W., Zukerman, I., Nicholson, A. E., and Bud, A. (1997).** "Towards a Bayesian model for keyhole plan recognition in large domains". In: *User Modeling*. Springer, pp. 365–376 (cited on pages 6, 76, 77).
- Alsallakh, B., Aigner, W., Miksch, S., and Groller, M. (2012).** "Reinventing the Contingency Wheel: Scalable Visual Analytics of Large Categorical Data". In: *Visualization and Computer Graphics, IEEE Transactions on* 18.12, pp. 2849–2858. ISSN: 1077-2626 (cited on page 94).
- Amabile, T. and Kramer, S. (2011).** *The progress principle: Using small wins to ignite joy, engagement, and creativity at work*. Harvard Business Press (cited on page 23).
- Amini, R. and Lisetti, C. (2013).** "HapFACS: An open source API/software to generate FACS-based expressions for ECAs animation and for corpus generation". In: *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, pp. 270–275 (cited on page 98).

- Anton, A., Earp, J. B., Young, J. D., et al. (2010).** "How internet users' privacy concerns have evolved since 2002". In: *Security & Privacy, IEEE 8.1*, pp. 21–27 (cited on page 132).
- Bakker, A. B., Schaufeli, W., and Van Dierendonck, D. (2000).** "Burnout: Prevalentie, risicogroepen en risicofactoren". In: *Psychische vermoeidheid en werk*, pp. 65–82 (cited on page 23).
- Bakker, J., Holenderski, L., Kocielnik, R., Pechenizkiy, M., and Sidorova, N. (2012).** "Stress work: From measuring stress to its understanding, prediction and handling with personalized coaching". In: *Proceedings of the 2<sup>nd</sup> ACM SIGHIT International health informatics symposium*. ACM, pp. 673–678 (cited on pages 12, 19, 30, 93, 106, 138).
- Baldauf, M., Dustdar, S., and Rosenberg, F. (2007).** "A survey on context-aware systems". In: *International Journal of Ad Hoc and Ubiquitous Computing 2.4*, pp. 263–277 (cited on page 140).
- Barkhuus, L. and Dey, A. K. (2003).** "Location-Based Services for Mobile Telephony: a Study of Users' Privacy Concerns." In: *INTERACT*. Vol. 3. Citeseer, pp. 702–712 (cited on page 168).
- Bernard, J., Ruppert, T., Scherer, M., Kohlhammer, J., and Schreck, T. (2012).** "Content-based Layouts for Exploratory Metadata Search in Scientific Research Data". In: *Proceedings of the 12<sup>th</sup> ACM/IEEE-CS Joint Conference on Digital Libraries*. JCDL '12. Washington, DC, USA: ACM, pp. 139–148. ISBN: 978-1-4503-1154-0 (cited on page 94).
- Bernard, J., Ruppert, T., Scherer, M., Schreck, T., and Kohlhammer, J. (2012).** "Guided Discovery of Interesting Relationships Between Time Series Clusters and Metadata Properties". In: *Proceedings of the 12<sup>th</sup> International Conference on Knowledge Management and Knowledge Technologies*. i-KNOW '12. Graz, Austria: ACM, 22:1–22:8. ISBN: 978-1-4503-1242-4 (cited on page 94).
- Bernard, J., Steiger, M., Widmer, S., Lücke-Tieke, H., May, T., and Kohlhammer, J. (2014).** "Visual-interactive Exploration of Interesting Multivariate Relations in Mixed Research Data Sets". In: *Computer Graphics Forum 33.3*, pp. 291–300 (cited on page 94).
- Bernard, J., Wilhelm, N., Kruger, B., May, T., Schreck, T., and Kohlhammer, J. (2013).** "Motionexplorer: Exploratory search in human motion capture data based on hierarchical aggregation". In: *Visualization and Computer Graphics, IEEE Transactions on 19.12*, pp. 2257–2266 (cited on page 94).

- Bodea, G., Huijboom, N., Kazemier, J., Koldijk, S., Koot, G., Munck, S. de, and Siljee, J. (2013). *Context-aware services: privacy concerns and strategies*. Tech. rep. The Netherlands: TNO (cited on page 172).
- Bokhove, W., Hulsebosch, B., Schoonhoven, B. v., Sappelli, M., and Wouters, K. (2012). "User privacy in applications for well-being and well-working. Requirements and approaches for user controlled privacy". In: *AMBIENT 2012: The second international conference on Ambient Computing, Applications, Services and Technologies*. (Cited on page 153).
- Borgo, R., Kehrner, J., Chung, D. H., Maguire, E., Laramée, R. S., Hauser, H., Ward, M., and Chen, M. (2013). "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications". In: *Eurographics State of the Art Reports*. EG STARs. Girona, Spain: Eurographics Association, pp. 39–63 (cited on page 94).
- Bosems, S. and van Sinderen, M. J. (2014). "Improving context-aware applications for the well-being domain: Model-driven design guided by medical knowledge". In: *Proceedings of the third International Conference on Sensor Networks, SENSORNETS 2014, Lisbon, Portugal*. Lisbon, Portugal: INSTICC, pp. 397–403 (cited on page 132).
- Bradley, M. M. and Lang, P. J. (1994). "Measuring emotion: the self-assessment manikin and the semantic differential". In: *Journal of behavior therapy and experimental psychiatry* 25.1, pp. 49–59 (cited on pages 65, 110).
- Bressert, S. (2015). *The Impact of Stress*. <http://psychcentral.com/lib/the-impact-of-stress/>. [Online; accessed: 2015-11-13. (Archived by WebCite® at <http://www.webcitation.org/6d0gSKCW0>)] (cited on page 19).
- Brightr (2015). <http://www.swell-project.net/results/swell-valorisation/brightr>. [Online; accessed: 2015-11-13. (Archived by WebCite® at <http://www.webcitation.org/6d0hqjQDJ>)] (cited on page 50).
- Caplan, R. D. (1987). "Person-environment fit theory and organizations: Commensurate dimensions, time perspectives, and mechanisms". In: *Journal of Vocational behavior* 31.3, pp. 248–267 (cited on page 26).
- Carroll, J. M. (2000). "Five reasons for scenario-based design". In: *Interacting with computers* 13.1, pp. 43–60 (cited on pages 132, 133, 143, 163).
- Caruana, R. and Niculescu-Mizil, A. (2006). "An empirical comparison of supervised learning algorithms". In: *Proceedings of the 23rd international conference on Machine learning*. ACM, pp. 161–168 (cited on page 112).



- Cavoukian, A. et al. (2009).** "Privacy by design: The 7 foundational principles". In: *Information and Privacy Commissioner of Ontario, Canada* (cited on page 147).
- Cavoukian, A. (2012).** "Operationalizing privacy by design: A guide to implementing strong privacy practices". In: *Information and Privacy Commissioner, Ontario, Canada* (cited on pages 168, 172).
- Choi, J. (2007).** "Context-driven requirements analysis". In: *Computational Science and Its Applications–ICCSA 2007*. Springer, pp. 739–748 (cited on pages 133, 137, 141, 162).
- Cinaz, B., Arnrich, B., La Marca, R., and Tröster, G. (2013).** "Monitoring of mental workload levels during an everyday life office-work scenario". In: *Personal and ubiquitous computing* 17.2, pp. 229–239 (cited on page 106).
- Commissioner, N. P. (2007).** *Privacy Impact Assessment Handbook* (cited on page 146).
- Compañó, R. and Lusoli, W. (2010).** "The policy maker's anguish: Regulating personal data behavior between paradoxes and dilemmas". In: *Economics of Information Security and Privacy*. Springer, pp. 169–185 (cited on pages 179, 187).
- Craig, S. D., D'Mello, S., Witherspoon, A., and Graesser, A. (2008).** "Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-affective states during learning". In: *Cognition and Emotion* 22.5, pp. 777–788 (cited on pages 30, 93, 107, 139).
- Czerwinski, M., Horvitz, E., and Wilhite, S. (2004a).** "A diary study of task switching and interruptions". In: *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. Vienna, Austria: ACM, pp. 175–182. ISBN: 1-58113-702-8 (cited on page 12).
- Czerwinski, M., Horvitz, E., and Wilhite, S. (2004b).** "A diary study of task switching and interruptions". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp. 175–182 (cited on page 76).
- Demerouti, E. and Bakker, A. B. (2011).** "The job demands-resources model: Challenges for future research". In: *SA Journal of Industrial Psychology* 37.2, pp. 01–09 (cited on pages 22, 23).
- Demerouti, E., Bakker, A. B., Geurts, S. A., and Taris, T. W. (2009).** "Daily recovery from work-related effort during non-work time". In: *Research in Occupational Stress and Well-being* 7, pp. 85–123 (cited on pages 24–26, 58).

- Demerouti, E., Bakker, A. B., Nachreiner, F., and Schaufeli, W. B. (2001). "The job demands-resources model of burnout". In: *Journal of Applied psychology* 86.3, p. 499 (cited on pages 12, 21, 22, 58, 131, 182).
- Dinev, T., Xu, H., Smith, J. H., and Hart, P. (2013). "Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts". In: *European Journal of Information Systems* 22.3, pp. 295–316 (cited on page 160).
- Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., Goldenstein, S. K., Vogler, C., Venkataraman, S., and Metaxas, D. N. (2005). "Optical computer recognition of facial expressions associated with stress induced by performance demands". In: *Aviation, space, and environmental medicine* 76.Supplement 1, B172–B182 (cited on pages 30, 61, 93, 107, 139).
- Doleisch, H., Gasser, M., and Hauser, H. (2003). "Interactive Feature Specification for Focus+Context Visualization of Complex Simulation Data". In: *Proceedings of the Symposium on Data Visualisation 2003. VISSYM '03*. Grenoble, France: Eurographics Association, pp. 239–248. ISBN: 1-58113-698-6 (cited on page 94).
- Drunen, A. van, Broek, E. L. van den, Spink, A. J., and Heffelaar, T. (2009). "Exploring workload and attention measurements with uLog mouse data". In: *Behavior research methods* 41.3, pp. 868–875 (cited on page 108).
- Duttweiler, P. C. (1984). "The internal control index: A newly developed measure of locus of control". In: *Educational and Psychological Measurement* 44.2, pp. 209–221 (cited on page 65).
- Epp, C., Lippold, M., and Mandryk, R. L. (2011). "Identifying emotional states using keystroke dynamics". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 715–724 (cited on page 108).
- Est, v. Q., Rerimassie, V. V., Keulen, v. I. I., and Dorren, G. G. (2014). *Intieme technologie: de slag om ons lichaam en gedrag*. Rathenau Instituut (cited on page 131).
- EU (1995). *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, OJ C L 281 (cited on page 147).
- EU (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of*

- personal data and on the free movement of such data*, OJ L 119 (cited on pages 179, 187).
- Fitbit (2015). <https://www.fitbit.com/nl>. [Online; accessed: 2015-11-13. (Archived by WebCite® at <http://www.webcitation.org/6d0hiHwLz>)] (cited on page 31).
- Flinn, S. and Lumsden, J. (2005). "User perceptions of privacy and security on the web". In: *Proceedings of 3<sup>rd</sup> Annual Conference on Privacy, Security and Trust*, pp. 15–26 (cited on page 132).
- Fogg, B. J. (2002). "Persuasive technology: using computers to change what we think and do". In: *Ubiquity* 2002.December, p. 5 (cited on page 34).
- French, J. R., Rodgers, W., Cobb, S., et al. (1974). "Adjustment as person-environment fit". In: *Coping and adaptation*, pp. 316–333 (cited on page 26).
- Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., and Roberts, J. C. (2011). "Visual comparison for information visualization". In: *Information Visualization* 10.4, pp. 289–309 (cited on page 94).
- Goalie (2015). <http://www.getgoalie.com/>. [Online; accessed: 2015-11-13. (Archived by WebCite® at <http://www.webcitation.org/6d0hV3vo7>)] (cited on page 31).
- Goldman, R. P., Geib, C. W., and Miller, C. A. (1999). "A new model of plan recognition". In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., pp. 245–254 (cited on page 77).
- Guyon, I. and Elisseeff, A. (2003). "An Introduction to Variable and Feature Selection". In: *J. Mach. Learn. Res.* 3, pp. 1157–1182. issn: 1532-4435 (cited on page 94).
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). "The WEKA data mining software: an update". In: *ACM SIGKDD explorations newsletter* 11.1, pp. 10–18 (cited on pages 79, 112).
- Harbers, M., Detweiler, C., and Neerincx, M. A. (2015). "Embedding Stakeholder Values in the Requirements Engineering Process". In: *Requirements Engineering: Foundation for Software Quality* 9013, pp. 318–332 (cited on page 160).
- Harbers, M., Aydogan, R., Jonker, C. M., and Neerincx, M. A. (2014). "Sharing information in teams: giving up privacy or compromising on team performance?" In: *Proceedings of the 2014 international conference on Autonomous*

- agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 413–420 (cited on page 152).
- Hart, S. G. and Staveland, L. E. (1988).** “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research”. In: *Advances in psychology* 52, pp. 139–183 (cited on pages 65, 110).
- Hartsuiker, M. (2015).** “Graphs versus avatars: Conveying data about stress in an analytical or an experiential way”. MA thesis. Leiden University (cited on page 41).
- Hassmen, P., Koivula, N., and Uutela, A. (2000).** “Physical exercise and psychological well-being: a population study in Finland”. In: *Preventive medicine* 30.1, pp. 17–25 (cited on page 25).
- Hoadley, C. M., Xu, H., Lee, J. J., and Rosson, M. B. (2010).** “Privacy as information access and illusory control: The case of the Facebook News Feed privacy outcry”. In: *Electronic commerce research and applications* 9.1, pp. 50–60 (cited on page 160).
- Hoepman, J.-H. (2014).** “Privacy design strategies”. In: *ICT Systems Security and Privacy Protection*. Springer, pp. 446–459 (cited on pages 134, 147, 149, 163, 172, 187).
- Hogervorst, M. A., Brouwer, A.-M., and Vos, W. K. (2013).** “Physiological correlates of stress in individuals about to undergo eye laser surgery”. In: *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, pp. 473–478 (cited on pages 30, 93, 138).
- Hong, J.-y., Suh, E.-h., and Kim, S.-J. (2009).** “Context-aware systems: A literature review and classification”. In: *Expert Systems with Applications* 36.4, pp. 8509–8522 (cited on page 130).
- Horizon2020 (2015).** <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/health-demographic-change-and-wellbeing/>. [Online; accessed: 2015-11-13.] (cited on page 1).
- Hulsebosch, B., Kooij, M., Kosman, R., and Janssen, J. (2013).** *D4.6 & D4.8 Design and evaluation of context privacy controls*. accessible at: <http://www.swell-project.net/results/deliverables> (cited on page 152).
- IJsselstein, W., Kort, Y. de, Midden, C., Eggen, B., and Hoven, E. van den (2006).** “Persuasive technology for human well-being: setting the scene”. In: *Persuasive technology*. Springer, pp. 1–5 (cited on pages 3, 13, 183).

- Ivancevich, J. M., Matteson, M. T., Freedman, S. M., and Phillips, J. S. (1990). "Worksite stress management interventions." In: *American Psychologist* 45.2, p. 252 (cited on pages 34, 182).
- Janssen, J. H., Broek, E. L. van den, and Westerink, J. H. (2012). "Tune in to your emotions: a robust personalized affective music player". In: *User Modeling and User-Adapted Interaction* 22.3, pp. 255–279 (cited on page 130).
- Johnson, S. M. and White, G. (1971). "Self-observation as an agent of behavioral change". In: *Behavior Therapy* 2.4, pp. 488–497 (cited on page 76).
- Kaasinen, E. (2003). "User needs for location-aware mobile services". In: *Personal and ubiquitous computing* 7.1, pp. 70–79 (cited on page 132).
- Kapoor, A. and Picard, R. W. (2005). "Multimodal affect recognition in learning environments". In: *Proceedings of the 13<sup>th</sup> annual ACM international conference on Multimedia*. ACM, pp. 677–682 (cited on pages 31, 61, 93, 107, 139).
- Karasek Jr, R. A. (1979). "Job demands, job decision latitude, and mental strain: Implications for job redesign". In: *Administrative science quarterly* 24.2, pp. 285–308 (cited on page 21).
- Kehrer, J. and Hauser, H. (2013). "Visualization and Visual Analysis of Multifaceted Scientific Data: A Survey". In: *Visualization and Computer Graphics, IEEE Transactions on* 19.3, pp. 495–513. issn: 1077-2626 (cited on page 93).
- Khan, I. A., Brinkman, W.-P., and Hierons, R. M. (2008). "Towards a Computer Interaction-Based Mood Measurement Instrument". In: *Proc. PPIG2008, ISBN*, pp. 971–978 (cited on page 108).
- Klasnja, P., Consolvo, S., Choudhury, T., Beckwith, R., and Hightower, J. (2009). "Exploring privacy concerns about personal sensing". In: *Pervasive Computing*. Springer, pp. 176–183 (cited on page 153).
- Knijnenburg, B. P. and Kobsa, A. (2013a). "Helping users with information disclosure decisions: potential for adaptation". In: *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, pp. 407–416 (cited on page 160).
- Knijnenburg, B. P. and Kobsa, A. (2013b). "Making decisions about privacy: information disclosure in context-aware recommender systems". In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 3.3, p. 20 (cited on pages 145, 159).
- Koldijk, S., Bernard, J., Ruppert, T., Kohlhammer, J., Neerincx, M., and Kraaij, W. (2015). "Visual Analytics of Work Behavior Data - Insights on Individual

- Differences". In: *Proceedings of Eurographics Conference on Visualization (EuroVis)* (cited on pages 91, 105, 123, 128).
- Koldijk, S., Kraaij, W., and Neerincx, M. A. (2016). "Deriving Requirements for Pervasive Well-Being Technology From Work Stress and Intervention Theory: Framework and Case Study". In: *Journal of Medical Internet Research: mHealth and uHealth* (cited on page 11).
- Koldijk, S., Neerincx, M. A., and Kraaij, W. (2012). "Unobtrusively measuring stress and workload of knowledge workers". In: *Proceedings of Measuring Behavior* (cited on page 89).
- Koldijk, S. (2011). "Look what you've done! Task recognition based on PC activities". MA thesis. Radboud University, Nijmegen, The Netherlands (cited on page 83).
- Koldijk, S. (2012). "Automatic recognition of context and stress to support knowledge workers". In: *Proceedings of the 30<sup>th</sup> European Conference on Cognitive Ergonomics* (cited on pages 12, 58, 104, 130, 169).
- Koldijk, S., Koot, G., Neerincx, M. A., and Kraaij, W. (2014). "Privacy and User Trust in Context-Aware Systems". In: *Proceedings of the 22<sup>nd</sup> Conference on User Modeling, Adaptation, and Personalization* (cited on pages 39, 42, 146, 147, 167).
- Koldijk, S., Sappelli, M., Neerincx, M. A., and Kraaij, W. (2013). "Unobtrusive monitoring of knowledge workers for stress self-regulation". In: *Proceedings of the 21<sup>st</sup> Conference on User Modeling, Adaptation, and Personalization* (cited on pages 30, 31, 71, 73, 94, 110, 139).
- Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M. A., and Kraaij, W. (2014). "The SWELL Knowledge Work Dataset for Stress and User Modeling Research". In: *Proceedings of the 16<sup>th</sup> International Conference on Multimodal Interaction* (cited on pages 57, 92, 109).
- Koldijk, S., Staalduinen, M. van, Neerincx, M. A., and Kraaij, W. (2012). "Real-time task recognition based on knowledge workers' computer activities". In: *Proceedings of the 30<sup>th</sup> European Conference on Cognitive Ergonomics* (cited on pages 30, 62, 75, 95, 137).
- Koldijk, S., Staalduinen, M. van, Raaijmakers, S., Rooij, I. van, and Kraaij, W. (2011). "Activity-logging for self-coaching of knowledge workers". In: *2nd workshop on information access for personal media archives* (cited on pages 86, 105).

- Koppes, L., Vroome, E. d., Mars, G., Janssen, B., Zwieten, M., and Bossche, S. v. d. (2012). *Nationale Enquete arbeidsomstandigheden 2012: Methodologie en globale resultaten* (cited on page 58).
- Korte, E. d., Blanson-Henkemans, O., Vliet, J. v., and Wiezer, N. (2014). *D1.5 Theoretical framework for behavioral change*. accessible at: <http://www.swell-project.net/results/deliverables> (cited on page 36).
- Kosara, R., Bendix, F., and Hauser, H. (2006). "Parallel Sets: Interactive Exploration and Visual Analysis of Categorical Data". In: *IEEE Transactions on Visualization and Computer Graphics* 12.4, pp. 558–568. issn: 1077-2626 (cited on page 94).
- Kraan, K., Dhondt, S., Houtman, I., Nelemans, R., and Vroome, E. d. (2000). *Handleiding NOVA-WEB: Hernieuwde versie*. (Cited on pages 2, 12, 27, 58).
- Laufer, R. S. and Wolfe, M. (1977). "Privacy as a concept and a social issue: A multidimensional developmental theory". In: *Journal of social Issues* 33.3, pp. 22–42 (cited on page 159).
- Le Fevre, M., Kolt, G. S., and Matheny, J. (2006). "Eustress, distress and their interpretation in primary and secondary occupational stress management interventions: which way first?" In: *Journal of Managerial Psychology* 21.6, pp. 547–565 (cited on pages 135, 136, 162).
- Le Fevre, M., Matheny, J., and Kolt, G. S. (2003). "Eustress, distress, and interpretation in occupational stress". In: *Journal of Managerial Psychology* 18.7, pp. 726–744 (cited on pages 17, 18).
- Leiter, M. P. and Maslach, C. (2003). "Areas of worklife: a structured approach to organizational predictors of job burnout". In: *Emotional and Physiological Processes and Positive Intervention Strategies*. Ed. by D. C. G. Pamela L. Perrewe. Emerald Group Publishing Limited. Chap. 4, pp. 91–134 (cited on page 26).
- Mark, G., Gudith, D., and Klocke, U. (2008). "The cost of interrupted work: more speed and stress". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, pp. 107–110 (cited on page 58).
- Maslach, C. and Leiter, M. P. (2008). "Early predictors of job burnout and engagement." In: *Journal of applied psychology* 93.3, p. 498 (cited on pages 16, 17, 181).
- Maslach, C., Schaufeli, W. B., and Leiter, M. P. (2001). "Job burnout". In: *Annual review of psychology* 52.1, pp. 397–422 (cited on page 17).



- Matthews, R., McDonald, N. J., and Trejo, L. J. (2005).** "Psycho-physiological sensor techniques: an overview". In: *11th International Conference on Human Computer Interaction (HCII)*, pp. 22–27 (cited on pages 104, 106, 186).
- May, T., Bannach, A., Davey, J., Ruppert, T., and Kohlhammer, J. (2011).** "Guiding feature subset selection with an interactive visualization". In: *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pp. 111–120 (cited on page 94).
- McDuff, D., El Kaliouby, R., Kassam, K., and Picard, R. (2011).** "Acume: A new visualization tool for understanding facial expression and gesture data". In: *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, pp. 591–596 (cited on page 98).
- Meijman, T. F., Mulder, G., Drenth, P., and Thierry, H. (1998).** "Psychological aspects of workload". In: *Handbook of Work and Organizational Psychology. Volume 2* (cited on pages 24, 25, 182).
- Mejean, L., Bicakova-Rocher, A., Kolopp, M., Villaume, C., Levi, F., Debry, G., Reinberg, A., and Drouin, P. (1988).** "Circadian and ultradian rhythms in blood glucose and plasma insulin of healthy adults". In: *Chronobiology international* 5.3, pp. 227–236 (cited on page 19).
- Michie, S. (2002).** "Causes and management of stress at work". In: *Occupational and Environmental Medicine* 59.1, pp. 67–72 (cited on pages 2, 12).
- Michie, S., Johnston, M., Francis, J., Hardeman, W., and Eccles, M. (2008).** "From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques". In: *Applied psychology* 57.4, pp. 660–680 (cited on pages 36, 182).
- Miller, L. (2015).** *Free to Fall*. HarperTeen (cited on page 190).
- Mokhayeri, F., Akbarzadeh-T, M.-R., and Toosizadeh, S. (2011).** "Mental stress detection using physiological signals based on soft computing techniques". In: *Biomedical Engineering (ICBME), 2011 18<sup>th</sup> Iranian Conference of*. IEEE, pp. 232–237 (cited on pages 60, 106).
- Nasoz, F., Lisetti, C. L., Alvarez, K., and Finkelstein, N. (2003).** "Emotion recognition from physiological signals for user modeling of affect". In: *Proceedings of the 3<sup>rd</sup> Workshop on Affective and Attitude User Modelling (Pittsburgh, PA, USA)* (cited on page 130).
- Natarajan, S., Bui, H. H., Tadepalli, P., Kersting, K., and Wong, W.-K. (2008).** "Logical hierarchical hidden markov models for modeling user activities".



- In: *Inductive Logic Programming*. Springer, pp. 192–209 (cited on pages 76, 88, 89, 185).
- Neerincx, M. A., Kennedie, S., Grootjen, M., and Grootjen, F. (2009). “Modeling the cognitive task load and performance of naval operators”. In: *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*. Springer, pp. 260–269 (cited on page 104).
- Neerincx, M. A. and Lindenberg, J. (2008). “Situating cognitive engineering for complex task environments”. In: *Naturalistic decision making and macrocognition*, pp. 373–390 (cited on pages 4, 13, 15, 29, 132, 162, 186).
- Norris, R., Carroll, D., and Cochrane, R. (1992). “The effects of physical activity and exercise training on psychological stress and well-being in an adolescent population”. In: *Journal of psychosomatic research* 36.1, pp. 55–65 (cited on page 25).
- Novak, D., Mihelj, M., and Munih, M. (2012). “A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing”. In: *Interacting with computers* 24.3, pp. 154–172 (cited on page 112).
- Office, U. I. C. (2014). *Conducting privacy impact assessments code of practice, version 1.0* (cited on page 146).
- Pantic, M., Pentland, A., Nijholt, A., and Huang, T. S. (2007). “Human computing and machine understanding of human behavior: a survey”. In: *Artificial Intelligence for Human Computing*. Springer, pp. 47–71 (cited on pages 107, 186).
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). “Context aware computing for the internet of things: A survey”. In: *Communications Surveys & Tutorials, IEEE* 16.1, pp. 414–454 (cited on pages 140, 141).
- Peters, G.-J. Y. (2015). *Intervention mapping*. <http://interventionmapping.com/?q=node/1711>. [Online; accessed: 2015-11-13. (Archived by WebCite® at <http://www.webcitation.org/6d0h2KQRL>)] (cited on page 34).
- Porayska-Pomsta, K., Anderson, K., Damian, I., Baur, T., André, E., Bernardini, S., and Rizzo, P. (2013). “Modelling users’ affect in job interviews: Technological demo”. In: *User Modeling, Adaptation, and Personalization*. Springer, pp. 353–355 (cited on page 130).
- Prekop, P. and Burnett, M. (2003). “Activities, context and ubiquitous computing”. In: *Computer Communications* 26.11, pp. 1168–1176 (cited on page 140).

- Rainie, L. (2005).** "Privacy Online: How Americans feel... the ways they are responding to new threats... and why they are changing their online behavior". In: *Pew Internet & American Life Project* (cited on page 160).
- Rath, A. S., Devaurs, D., and Lindstaedt, S. N. (2009).** "UICO: an ontology-based user interaction context model for automatic task detection on the computer desktop". In: *CIAO '09: Proceedings of the 1<sup>st</sup> Workshop on Context, Information and Ontologies*. Heraklion, Greece: ACM, pp. 1–10. ISBN: 978-1-60558-528-4 (cited on pages 6, 76).
- Richardson, K. M. and Rothstein, H. R. (2008).** "Effects of occupational stress management intervention programs: a meta-analysis." In: *Journal of occupational health psychology* 13.1, p. 69 (cited on page 34).
- Richman, G. S., Riordan, M. R., Reiss, M. L., Pyles, D. A., and Bailey, J. S. (1988).** "The effects of self-monitoring and supervisor feedback on staff performance in a residential setting." In: *Journal of Applied Behavior Analysis* 21.4, p. 401 (cited on page 76).
- Riera, A., Soria-Frisch, A., Albajes-Eizagirre, A., Cipresso, P., Grau, C., Dunne, S., Ruffini, G., and Barcelona, S. aStarlab (2012).** "Electro-physiological data fusion for stress detection". In: *Studies in health technology and informatics* 181, pp. 228–32 (cited on pages 60, 106).
- Romanosky, S., Acquisti, A., Hong, J., Cranor, L. F., and Friedman, B. (2006).** "Privacy patterns for online interactions". In: *Proceedings of the 2006 conference on Pattern languages of programs*. ACM, p. 12 (cited on page 147).
- Rossi, E. L. and Nimmons, D. (1991).** *The 20-minute break: reduce stress, maximize performance, and improve health and emotional well-being using the new science of ultradian rhythms*. JP Tarcher (cited on page 19).
- Rubinstein, I. (2012).** "Big data: the end of privacy or a new beginning?" In: *International Data Privacy Law (2013 Forthcoming)*, pp. 12–56 (cited on pages 131, 179).
- Ruff, J. (2002).** "Information Overload: Causes, Symptoms and Solutions". In: *Harvard Graduate School of Education*, pp. 1–13 (cited on page 58).
- Salmeron-Majadas, S., Santos, O. C., and Boticario, J. (2013).** "Inclusive Personalized e-Learning Based on Affective Adaptive Support." In: *UMAP*. Springer, pp. 384–387 (cited on page 130).
- Sappelli, M. and Verberne, S. (2014).** *D3.7 Context and task sensitive information filtering and sharing agent*. accessible at: <http://www.swell-project.net/results/deliverables> (cited on page 48).

- Sappelli, M. (2016).** "Knowledge work in context: User centered knowledge worker support". PhD thesis. Radboud University Nijmegen (cited on pages 30, 48, 137).
- Sappelli, M., Verberne, S., Koldijk, S., and Kraaij, W. (2014).** "Collecting a dataset of information behaviour in context". In: *Proceedings of the 4<sup>th</sup> Workshop on Context-Awareness in Retrieval and Recommendation*. ACM, pp. 26–29 (cited on pages 49, 60, 73, 92).
- Schaufeli, W. B., Salanova, M., González-Romá, V., and Bakker, A. B. (2002).** "The measurement of engagement and burnout: A two sample confirmatory factor analytic approach". In: *Journal of Happiness studies* 3.1, pp. 71–92 (cited on page 17).
- Schavemaker, J., Koldijk, S., and Boertjes, E. (2014).** *D3V1.1 Fishualization release and documents*. accessible at: <http://www.swell-project.net/results/deliverables> (cited on page 42).
- Schavemaker, J., Boertjes, E., Koldijk, S., Wiertz, L., Verberne, S., Sappelli, M., and Kaptein, R. (2014).** "Fishualization: a group feedback display". In: *Measuring Behavior 2014 - 9<sup>th</sup> International Conference on Methods and Techniques in Behavioral Research* (cited on page 42).
- Schendel, J. A. v. (2015).** "A Matter of Time: The Influence of Context-Based Timing on Compliance with Well-Being Triggers". MA thesis. Eindhoven University of Technology (cited on page 48).
- Scherer, S., Glodek, M., Layher, G., Schels, M., Schmidt, M., Brosch, T., Tschechne, S., Schwenker, F., Neumann, H., and Palm, G. (2012).** "A generic framework for the inference of user states in human computer interaction". In: *Journal on Multimodal User Interfaces* 6.3-4, pp. 117–141 (cited on page 128).
- Selye, H. (1950).** "Stress and the general adaptation syndrome". In: *British medical journal* 1.4667, p. 1383 (cited on page 19).
- Selye, H. (1956).** *The stress of life*. McGraw-Hill (cited on pages 12, 181).
- Selye, H. (1975).** "Stress without distress." In: *Vie medicale au Canada francais* (cited on page 18).
- Setz, C., Arnrich, B., Schumm, J., La Marca, R., Troster, G., and Ehlert, U. (2010).** "Discriminating stress from cognitive load using a wearable EDA device". In: *Information Technology in Biomedicine, IEEE Transactions on* 14.2, pp. 410–417 (cited on page 106).

- Seyff, N., Graf, F., Grünbacher, P., and Maiden, N. (2008). "Mobile discovery of requirements for context-aware systems". In: *Requirements Engineering: Foundation for Software Quality*. Springer, pp. 183–197 (cited on page 133).
- Sharma, N. and Gedeon, T. (2012). "Objective measures, sensors and computational techniques for stress recognition and classification: A survey". In: *Computer methods and programs in biomedicine* 108.3, pp. 1287–1301 (cited on pages 7, 30, 93, 138).
- Shen, J., Brdiczka, O., and Liu, J. (2013). "Understanding email writers: Personality prediction from email messages". In: *User Modeling, Adaptation, and Personalization*. Springer, pp. 318–330 (cited on page 130).
- Shin, D.-H. (2010). "The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption". In: *Interacting with Computers* 22.5, pp. 428–438 (cited on pages 168, 176, 177).
- Siegrist, J. (2012). "Effort-reward imbalance at work: Theory, measurement and evidence". In: *Department of Medical Sociology, University Düsseldorf, Düsseldorf* (cited on page 24).
- Smith, H. J., Dinev, T., and Xu, H. (2011). "Information privacy research: an interdisciplinary review". In: *MIS quarterly* 35.4, pp. 989–1016 (cited on pages 132, 160, 168, 179, 180).
- Smith, H. J., Milberg, S. J., and Burke, S. J. (1996). "Information privacy: measuring individuals' concerns about organizational practices". In: *MIS quarterly*, pp. 167–196 (cited on page 132).
- Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). "A multimodal database for affect recognition and implicit tagging". In: *Affective Computing, IEEE Transactions on* 3.1, pp. 42–55 (cited on pages 61, 107, 184).
- Solove, D. J. (2006). "A taxonomy of privacy". In: *University of Pennsylvania law review*, pp. 477–564 (cited on page 147).
- Sonnentag, S. and Fritz, C. (2007). "The Recovery Experience Questionnaire: development and validation of a measure for assessing recuperation and unwinding from work." In: *Journal of occupational health psychology* 12.3, p. 204 (cited on page 24).
- Spiekermann, S., Grossklags, J., and Berendt, B. (2001). "E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior". In: *Proceedings of the 3<sup>rd</sup> ACM conference on Electronic Commerce*. ACM, pp. 38–47 (cited on page 132).

- Stolte, C. and Hanrahan, P. (2000).** "Polaris: A System for Query, Analysis and Visualization of Multi-Dimensional Relational Databases". In: *Proceedings of the IEEE Symposium on Information Visualization 2000*. INFOVIS '00. Washington, DC, USA: IEEE Computer Society, pp. 5–. ISBN: 0-7695-0804-9 (cited on page 94).
- Sutcliffe, A. G. (2013).** "Requirements engineering". In: *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* (Cited on pages 132, 134, 151, 163).
- SWELL project (2015).** <http://www.swell-project.net/>. [Online; accessed: 2015-11-13. (Archived by WebCite® at <http://www.webcitation.org/6d0hzJXbt>)] (cited on page 12).
- Tahboub, K. A. (2006).** "Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition". In: *Journal of Intelligent and Robotic Systems* 45.1, pp. 31–52 (cited on page 88).
- Tao, J. and Tan, T. (2005).** "Affective computing: A review". In: *Affective computing and intelligent interaction*. Springer, pp. 981–995 (cited on page 130).
- Toch, E., Wang, Y., and Cranor, L. F. (2012).** "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems". In: *User Modeling and User-Adapted Interaction* 22.1-2, pp. 203–220 (cited on page 149).
- Van De Garde-Perik, E., Markopoulos, P., De Ruyter, B., Eggen, B., and Ijsselstein, W. (2008).** "Investigating privacy attitudes and behavior in relation to personalization". In: *Social Science Computer Review* 26.1, pp. 20–43 (cited on page 131).
- Vizer, L. M., Zhou, L., and Sears, A. (2009).** "Automated stress detection using keystroke and linguistic features: An exploratory study". In: *International Journal of Human-Computer Studies* 67.10, pp. 870–886 (cited on pages 31, 61, 93, 108, 139, 186).
- Vos, P., Staaldin, M. van, and Koldijk, S. (2010).** *Studie naar werkvreugde en determinanten: Literatuurstudie en reflectie van experts*. Tech. rep. The Netherlands: TNO (cited on pages 16, 34).
- Wabeke, T. (2014).** "Recommending tips that support wellbeing at work to knowledge workers". MA thesis. Radboud University, Nijmegen, The Netherlands (cited on pages 45, 136, 153, 154).
- Wright, D. (2012).** "The state of the art in privacy impact assessment". In: *Computer Law & Security Review* 28.1, pp. 54–61 (cited on pages 134, 146, 163, 170, 187).

- Xu, H. (2007).** "The effects of self-construal and perceived control on privacy concerns". In: *ICIS 2007 Proceedings*, p. 125 (cited on page 160).
- Yerkes, R. M. and Dodson, J. D. (1908).** "The relation of strength of stimulus to rapidity of habit-formation". In: *Journal of Comparative Neurology and Psychology* 18.5, pp. 459–482 (cited on page 18).
- Zapf, D. (1993).** "Stress-oriented analysis of computerized office work". In: *The European Work and Organizational Psychologist* 3.2, pp. 85–100 (cited on pages 60, 104).
- Zijlstra, F. and Doorn, L. van (1985).** "The construction of a scale to measure subjective effort". PhD thesis. Delft University of Technology (cited on pages 65, 110).
- Zwieten, M. v., Vroome, E. d., Mol, M., Mars, M., Koppes, L., and Bossche, S. v. d. (2014).** *Nationale Enquête Arbeidsomstandigheden 2013. Methodologie en globale resultaten*. Tech. rep. The Netherlands: TNO (cited on pages 2, 12, 27).



## Summary

Many people experience stress at work. In this thesis, we investigated how new technologies, for example sensors and apps, can be used to help people cope better with stress.

To start, we looked at several relevant **work stress theories**. We found that the term ‘stress’ is used for different things. At the workplace there can be different stressors (e.g. task demands, interruptions, time pressure). These can, dependent on the characteristics of the person, lead to the perception of stress. This can cause acute stress reactions, e.g. a change in heart rhythm. Without relaxation and recovery, stress can build up and cause long-term effects, such as head aches, being irritated, and in the worst case burn-out (which is characterized by exhaustion, cynicism and low self-efficacy). There are no general causes that always lead to stress. Work is experienced as stressful, when there is a disbalance between high task demands on the one side and on the other side: too little resources, too few rewards, too little recovery, or an environment that does not fit the person.

We investigated **how new technologies can be used** to improve well-being at work. We figured out that sensors could be used to measure aspects of the work itself (e.g. the tasks and topics worked on), to measure acute stress (e.g. heart rate variability), or to measure long-term effects of stress and recovery (e.g. sleep or amount of physical activity). Computer systems could be used to address stressors (e.g. providing support during work), to address acute stress (e.g. helping to better cope with stress), or to address long-term stress consequences (e.g. helping to relax and recover). Ideally, the system uses behavioral change techniques to motivate the user, improve his skills, or provide a reminder.

Moreover, we investigated how we can use **sensors in the office** to measure stress and the context in which it appears. To start, we set up an experiment, in which 25 people were asked to write reports and make presentations, while we manipulated their working conditions with stressors: in one condition they worked as they would usually do, in one condition we sent them many emails,



and in one condition we gave them time pressure. After each condition, we let the participants fill in questionnaires to assess how stressed they were, what their task load was, how high their mental effort was, and what their emotion was. While they were working, we collected data with several sensors. Computer interactions were recorded with a computer logger, facial expressions were recorded with a web-cam, body postures were recorded with a Kinect, and physiology (heart rate and skin conductance as signs of stress) was recorded with body sensors.

Then, we investigated whether we can use computer interactions to estimate **which task** someone is performing (e.g. writing a report, making an overview, making a planning). We found that people work very differently. So it is best to make a model per person. With only 30 examples (ca. 2.5 hours of working data), we were able to make a model that can recognize 12 different tasks with an accuracy of 80%.

Moreover, we investigated whether we can use sensor data to provide insight in **work behavior, related to stress**. Therefore, we combined data visualization techniques with data analysis (called 'Visual Analytics'). We found a relation between mental effort and facial expressions. Therefore, we made a visualization to show patterns in facial expressions over time, and added an avatar. There were individual differences regarding the way someone exactly looked. We also investigated which facial expressions are typically associated with a high and low mental effort. Using data visualization methods gave us many new insights into our dataset.

Finally, we investigated whether we can make an automatic **stress estimate**. First of all, we tested whether we can determine if a data point is from a neutral working condition, or a working condition with stressors. Our model was able to distinguish the working conditions with an accuracy of 90%. Body postures gave the most useful information, followed by facial expressions. Secondly, we investigated which subjective variable can best be predicted. We found that, based upon sensor data, mental effort can be estimated with an accuracy (correlation) of 0.84. Facial expressions gave the most useful information, followed by postures. Finally, we investigated in how far individual differences play a role. Especially for predicting mental effort, models benefit from information about the participant. We trained models on subgroups of users that are similar, and those models almost always outperformed general models.

In our research, we also paid attention to the **end user**. We extended an existing development method, to find all requirements for a context aware support system. A good design is based upon insights from theories, as well as possibilities of the technology, and requirements from the end user. We especially focused on possible concerns of users, in particular regarding privacy, and how

---

these can be addressed in the design. Besides asking the users themselves, we performed a so called Privacy Impact Assessment, to find all privacy issues that play a role in our system. Furthermore, we looked at Privacy by Design, in which privacy-friendly choices, e.g. regarding data collection, are already made in the design process.

As a last thing, we evaluated with users, what the effect of **Privacy by Design** is on trust in the system, and the intention to use it. The people that got information on Privacy by Design had less privacy concerns and more trust in the system. The intention to use the system, however, depended mainly on how useful they found it, and was unrelated to privacy concerns. This ‘privacy paradox’ was also found in previous research. To address the user’s privacy, it is therefore important to apply privacy by design.

All in all we can say that in this research we successfully combined 3 perspectives - work stress theory, sensor technology and user needs - to find ways to support employees to cope better with stress.



# Samenvatting

Veel mensen hebben last van stress op het werk. In dit proefschrift hebben we onderzocht hoe nieuwe technologieën, bv. sensoren en apps, ingezet kunnen worden om mensen te helpen beter met stress om te gaan.

Om te beginnen, hebben we gekeken naar relevante **werkstress theorieën**. Hieruit blijkt dat de term 'stress' voor verschillende dingen wordt gebruikt. Op het werk kunnen er verschillende stressoren zijn (bv. taakeisen, interrupties, tijdsdruk). Deze kunnen, afhankelijk van de persoon, leiden tot de perceptie van stress. Dit kan zorgen voor acute stressverschijnselen, bv. een verandering van het hartritme. Zonder ontspanning en herstel kan stress opbouwen en leiden tot lange termijn verschijnselen, zoals hoofdpijn, geïrriteerd zijn en in het ergste geval burn-out (gekaracteriseerd door uitputting, cynisme en weinig vertrouwen in het eigen kunnen). Er zijn geen generieke oorzaken die altijd tot stress leiden. Werk wordt als stressvol ervaren als er een disbalans is tussen hoge taakeisen aan de ene kant en aan de andere kant: te weinig regelmogelijkheden, te weinig waardering, te weinig herstel of een omgeving die niet past bij de persoon.

We hebben gekeken naar **hoe nieuwe technologieën ingezet kunnen worden** om het welzijn op het werk te verbeteren. We hebben gevonden dat sensoren gebruikt zouden kunnen worden om eigenschappen van het werk zelf te meten (bv. het soort taken en onderwerpen), om acute stress te meten (bv. hartslag variabiliteit) of om lange termijneffecten van stress en herstel te meten (bv. slaap of mate van fysieke activiteit). Computersystemen zouden kunnen ingezet worden om stressoren aan te pakken (bv. ondersteuning tijdens het werk bieden), acute stress aan te pakken (b.v. helpen beter met stress om te gaan) of lange termijn stress-consequenties aan te pakken (bv. helpen bij ontspanning en herstel). Idealiter gebruikt het systeem gedragsveranderingstechnieken om de gebruiker te motiveren zijn vaardigheden te verbeteren of een herinnering te geven.

Verder hebben we gekeken hoe **sensoren op de werkplek** ingezet kunnen

worden om stress en de context waarin stress optreedt te meten. We hebben om te beginnen een experiment opgezet, waarin we 25 mensen hebben gevraagd om verslagen te schrijven en presentaties te maken, terwijl wij de werkomstandigheden hebben gemanipuleerd met stressoren: we hebben ze in een experimentele conditie gewoon laten werken, in een andere conditie hebben we ze heel veel e-mails gestuurd en in een conditie hebben we ze onder tijdsdruk gezet. Na elke conditie hebben we de deelnemers d.m.v. vragenlijsten gevraagd, hoe gestrest ze waren, wat hun taaklast was, hoe hoog hun mentale inspanning was en wat hun emotie was. Terwijl ze aan het werk waren, hebben we data verzameld d.m.v. verschillende sensoren. Computerinteracties hebben we opgenomen met een computer logger, gezichtsuitdrukkingen hebben we opgenomen met een webcam, lichaamshoudingen hebben we opgenomen met een Kinect, en fysiologie (hartslag en huidgeleiding als teken van stress) hebben we opgenomen met lichaamssensoren.

Vervolgens hebben we onderzocht of we op basis van computerinteracties een inschatting kunnen maken van **welke taak** iemand aan het uitvoeren is (bv. een verslag schrijven, een overzicht maken, een planning maken). We kwamen erachter dat mensen erg verschillend werken. Het is dus het beste om per persoon een model te maken. We waren in staat om met maar 30 voorbeelden (ca. 2.5 uur werkdata) een model te maken dat met 80% nauwkeurigheid 12 verschillende taken kan onderscheiden.

Verder hebben we onderzocht of we met sensordata inzicht kunnen bieden in **werkgedrag, gerelateerd aan stress**. Hiervoor hebben we data visualisatie methoden gecombineerd met data analyse ('Visual Analytics' genoemd). We hebben een samenhang gevonden tussen mentale inspanning en gezichtsuitdrukkingen. Daarom hebben we een visualisatie gemaakt om patronen in gezichtsuitdrukkingen over de tijd te visualiseren, en een avatar toegevoegd om specifieke gezichtsuitdrukkingen te tonen. Er waren individuele verschillen in hoe iemand precies kijkt. We hebben ook onderzocht welke typische gezichtsuitdrukkingen voorkwamen bij een hoge en lage mentale werklast. Het gebruiken van data visualisatie methoden heeft ons veel nieuwe inzichten gegeven in de dataset.

We hebben onderzocht of we automatisch een **stress-inschatting** kunnen maken. Eerst hebben we gekeken of we kunnen herkennen dat een datapunt uit de gewone werkconditie komt, of uit een werkconditie met stressoren. Ons model kon de werkcondities met 90% nauwkeurigheid onderscheiden. Lichaamshouding gaf de meeste informatie, gevolgd door gezichtsuitdrukkingen. Daarna hebben we gekeken welke subjectieve variabele we het best kunnen voorspellen. Het bleek dat op basis van sensordata de mentale inspanning met een nauwkeurigheid (correlatie) van .84 voorspeld kan

---

worden. Gezichtsuitdrukkingen geven hier de meeste informatie, gevolgd door lichaamshouding. Ook hebben we gekeken in hoeverre individuele verschillen een rol spelen. Vooral bij het voorspellen van mentale inspanning blijken modellen veel baat te hebben bij informatie over de proefpersoon. We hebben modellen getraind op subgroepen van gebruikers die op elkaar leken, en die modellen werkten eigenlijk altijd beter dan de generieke modellen.

In ons onderzoek hebben we ook aandacht besteed aan de **eindgebruiker**. We hebben een bestaande ontwikkelmethode uitgebreid om op een systematische manier alle eisen aan het systeem in kaart te brengen. Een goed ontwerp neemt zowel inzichten vanuit de theorie mee, als ook de mogelijkheden van de technologie en eisen vanuit de eindgebruiker. We hebben vooral gekeken naar mogelijke bezwaren van de eindgebruiker, vooral rondom privacy, en hoe we die konden plaatsen in het ontwerp. Behalve de gebruikers zelf te vragen naar bezwaren, hebben we een zogenaamd Privacy Impact Assessment uitgevoerd, om de privacy issues in ons systeem in kaart te brengen. Verder hebben we gekeken naar Privacy by Design, waarbij in het systeemontwerp al privacy-vriendelijke keuzes werden gemaakt, bv. rondom datacollectie.

Als laatste hebben we met gebruikers geëvalueerd, wat het effect van informatie over **Privacy by Design** is op het vertrouwen in het systeem en het gebruik van het systeem. De mensen die informatie over Privacy by Design kregen, hadden minder privacy-bezwaren en meer vertrouwen in het systeem. De intentie om het systeem te gebruiken was echter vooral afhankelijk van hoe nuttig mensen het systeem vonden en stond los van privacy-bezwaren. Deze 'privacy-paradox' is eerder gevonden in ander onderzoek. Om de privacy van de gebruiker te beschermen, is het daarom belangrijk om privacy by design toe te passen in het ontwerp.

Samenvattend kunnen we zeggen dat we in dit onderzoek 3 perspectieven succesvol gecombineerd hebben - werkstress theorie, sensor technologie en gebruikerseisen - om manieren te vinden om werknemers te ondersteunen beter met stress om te gaan.



# Zusammenfassung

Viele Menschen erleben Stress am Arbeitsplatz. In dieser Doktorarbeit haben wir untersucht, wie neue Technologien, z.B. Sensoren und Apps, verwendet werden können, um Menschen zu helfen besser mit Stress umzugehen.

Als erstes haben wir uns relevante **Arbeitsstresstheorien** angeschaut. Wir haben festgestellt, dass der Begriff 'Stress' für verschiedene Dinge verwendet wird. Am Arbeitsplatz können verschiedene Stressoren auftreten z.B. Arbeitsanforderungen, Unterbrechungen, Zeitdruck. Diese können, abhängig von der Person, zu einer Wahrnehmung von Stress führen. Dies kann zu akuten Stresssymptomen führen, z.B. eine Veränderung des Herzrhythmus. Ohne Entspannung und Erholung kann Stress sich anhäufen und Langzeitfolgen haben, wie Kopfschmerzen, Gereiztheit sein, und im schlimmsten Fall Burn-out (was charakterisiert wird durch Erschöpfung, Zynismus und wenig Vertrauen in das eigene Können). Es gibt keine generellen Ursachen, die immer zu Stress führen. Arbeit wird vor allem als stressig erfahren, wenn es eine Unausgewogenheit gibt zwischen hohen Arbeitsanforderungen auf der einen Seite und auf der anderen Seite: Zu wenig Ressourcen, zu wenig Belohnung, zu wenig Erholung oder eine Umgebung, die nicht zur Person passt.

Wir haben untersucht, **wie neue Technologien eingesetzt werden können**, um das Wohlbefinden am Arbeitsplatz zu verbessern. Wir haben herausgefunden, dass Sensoren eingesetzt werden könnten, um Eigenschaften der Arbeit selbst zu messen (z.B. Art oder Inhalt der Tätigkeit), um akuten Stress zu messen (z.B. Herzschlagvariabilität) oder um Langzeiteffekte von Stress und Erholung zu messen (z.B. Schlaf oder Ausmaß an körperlicher Aktivität). Unterstützende Computersysteme könnten sich auf Stressoren richten (z.B. Unterstützung während der Arbeit bieten), auf akuten Stress (z.B. helfen besser mit Stress umzugehen) oder auf Langzeitkonsequenzen (z.B. helfen bei Entspannung und Erholung). Im Idealfall benutzt das System Verhaltensveränderungstechniken, um den Benutzer zu motivieren, seine Fähigkeiten zu verbessern oder eine Erinnerung zu bieten.



Außerdem haben wir geschaut, wie **Sensoren am Arbeitsplatz** eingesetzt werden können, um Stress und den Kontext in dem er auftritt, zu messen. Wir haben zuerst in einer Studie 25 Menschen gebeten, Berichte zu schreiben und Präsentationen vorzubereiten, während wir ihre Arbeitsbedingungen manipuliert haben mit Stressoren: Wir haben sie in einer Experimentalbedingung arbeiten lassen wie gewöhnlich, in einer Experimentalbedingung haben wir ihnen viele Emails geschickt, und in einer Experimentalbedingung haben wir sie unter Zeitdruck gesetzt. Nach jeder Experimentalbedingung haben wir die Teilnehmer mittels Fragebögen gefragt, wie gestresst sie waren, was ihre Arbeitsbelastung war, wie hoch ihre mentale Anstrengung war und was ihre Emotion war. Während sie gearbeitet haben, haben wir Daten gesammelt mittels verschiedener Sensoren. Computer Interaktionen haben wir aufgezeichnet mit einer Computerlogging Software, Mimik haben wir aufgezeichnet mit einer Webcam, Körperhaltung haben wir aufgezeichnet mit einer Kinect und Physiologie (Herzschlag und Hautleitfähigkeit als Anzeichen von Stress) haben wir aufgezeichnet mit Körpersensoren.

Als nächstes haben wir untersucht, ob wir basierend auf Computerinteraktionen einschätzen können, **welche Art der Tätigkeit** jemand gerade ausführt (z.B. einen Bericht schreiben, eine Übersicht machen, einen Zeitplan machen). Wir haben herausgefunden, dass Menschen sehr verschieden arbeiten. Es ist also am besten, pro Person ein Modell zu machen. Wir waren in der Lage mit nur 30 Beispielen (2.5 Stunden Arbeitsdaten) ein Modell zu machen, dass mit 80% Genauigkeit 12 verschiedene Tätigkeiten unterscheiden kann.

Außerdem haben wir untersucht, ob wir an Hand von Sensordaten Einsicht geben können in **Arbeitsverhalten im Zusammenhang mit Stress**. Dazu haben wir Methoden der Datenvisualisierung kombiniert mit Datenanalyse ('Visual Analytics' genannt). Wir haben einen Zusammenhang gefunden zwischen mentaler Anstrengung und Mimik. Daher haben wir eine Visualisierung gemacht, um Muster in der Mimik im Verlauf der Zeit zu visualisieren und haben einen Avatar hinzugefügt. Es gab individuelle Unterschiede in der genauen Mimik. Wir haben außerdem untersucht, welche Mimik typisch ist für eine hohe und niedrige mentale Anstrengung. Das Anwenden von Methoden der Datenvisualisierung hat uns viele neue Einsichten in unseren Datensatz gegeben.

Weiterhin haben wir untersucht, ob wir automatisch eine **Stresseinschätzung** machen können. Als erstes haben wir geschaut, ob wir erkennen können, ob ein Datenpunkt aus der normale Experimentalbedingung stammt oder aus der Experimentalbedingung mit Stressoren. Unser Modell konnte mit

---

90% Genauigkeit die Experimentalbedingungen unterscheiden. Körperhaltung war am informativsten, gefolgt von Mimik. Zweitens haben wir geschaut, welche subjektive Variable wir am besten vorhersagen können. Wir haben herausgefunden, dass basierend auf Sensordaten, die mentale Anstrengung mit einer Genauigkeit (Korrelation) von .84 eingeschätzt werden kann. Mimik war am informativsten, gefolgt von Körperhaltung. Schließlich haben wir geschaut, inwiefern individuelle Unterschiede eine Rolle spielen. Vor allem beim Einschätzen der mentalen Anstrengung erwiesen sich Informationen über die Versuchsperson als nützlich. Wir haben Modelle für Teilgruppen der Benutzer, die einander ähnlich sind, gemacht und diese Modelle funktionierten eigentlich immer besser als generische Modelle.

In unserer Forschung haben wir auch den **Endnutzer** miteinbezogen. Wir haben eine bestehende Entwicklungsmethode erweitert, um systematisch alle Anforderungen an das System zu finden. Ein guter Systementwurf verbindet theoretische Einsichten mit technologischen Möglichkeiten und Anforderungen von Endnutzern. Wir haben uns konzentriert auf mögliche Bedenken von Endnutzern, vor allem im Bereich von Privatsphäre und Datenschutz und die Konsequenzen für den Systementwurf. Wir haben nicht nur die Endnutzer selbst gefragt, sondern auch ein sogenanntes Privacy Impact Assessment durchgeführt, um alle Privatsphäre- und Datenschutz-Aspekte, die eine Rolle spielen in unserem System zu finden. Außerdem haben wir uns Privacy by Design angeschaut, wobei schon im Systementwurf privatsphären-freundliche Entscheidungen getroffen werden, z.B. über die Datensammlung.

Zuletzt haben wir mit Endnutzern evaluiert, was der Effekt von Informationen über **Privacy by Design** ist, auf das Vertrauen ins System und die Intention das System zu nutzen. Die Leute, die Informationen über Privacy by Design bekommen haben, hatten weniger Bedenken hinsichtlich der Privatsphäre und mehr Vertrauen ins System. Die Intention, das System zu nutzen, war aber hauptsächlich abhängig davon, wie nützlich sie das System fanden und unabhängig von Bedenken hinsichtlich der Privatsphäre. Dieses 'privacy paradox' ist auch in vorgehenden Forschungen gefunden worden. Um die Privatsphäre der Endnutzer zu schützen, ist es daher wichtig, Privacy by Design beim Entwurf anzuwenden.

Zusammenfassend können wir sagen, dass wir in dieser Forschungsarbeit 3 Perspektiven erfolgreich kombiniert haben - Arbeitsstresstheorie, Sensortechnologie und Nutzeranforderungen - um Wege zu finden, Mitarbeiter zu unterstützen, Stress besser zu bewältigen.



# Acknowledgements

Thanks.

## SIKS Dissertatiereeks

---

2009

---

- 2009-01 Rasa Jurgelenaite (RUN)  
Symmetric Causal Independence Models
- 2009-02 Willem Robert van Hage (VU)  
Evaluating Ontology-Alignment Techniques
- 2009-03 Hans Stol (UvT)  
A Framework for Evidence-based Policy Making Using IT
- 2009-04 Josephine Nabukenya (RUN)  
Improving the Quality of Organisational Policy Making using Collaboration Engineering
- 2009-05 Sietse Overbeek (RUN)  
Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality
- 2009-06 Muhammad Subianto (UU)  
Understanding Classification
- 2009-07 Ronald Poppe (UT)  
Discriminative Vision-Based Recovery and Recognition of Human Motion
- 2009-08 Volker Nannen (VU)  
Evolutionary Agent-Based Policy Analysis in Dynamic Environments
- 2009-09 Benjamin Kanagwa (RUN)  
Design, Discovery and Construction of Service-oriented Systems
- 2009-10 Jan Wielemaker (UVA)  
Logic programming for knowledge-intensive interactive applications
- 2009-11 Alexander Boer (UVA)  
Legal Theory, Sources of Law & the Semantic Web
- 2009-12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin)  
Operating Guidelines for Services
- 2009-13 Steven de Jong (UM)  
Fairness in Multi-Agent Systems
- 2009-14 Maksym Korotkiy (VU)  
From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)
- 2009-15 Rinke Hoekstra (UVA)  
Ontology Representation - Design Patterns and Ontologies that Make Sense
- 2009-16 Fritz Reul (UvT)  
New Architectures in Computer Chess
- 2009-17 Laurens van der Maaten (UvT)  
Feature Extraction from Visual Data
- 2009-18 Fabian Groffen (CWI)  
Armada, An Evolving Database System
- 2009-19 Valentin Robu (CWI)  
Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets
- 2009-20 Bob van der Vecht (UU)  
Adjustable Autonomy: Controlling Influences on Decision Making
- 2009-21 Stijn Vanderlooy (UM)  
Ranking and Reliable Classification
- 2009-22 Pavel Serdyukov (UT)  
Search For Expertise: Going beyond direct evidence
- 2009-23 Peter Hofgesang (VU)  
Modelling Web Usage in a Changing Environment
- 2009-24 Annerieke Heuvelink (VUA)  
Cognitive Models for Training Simulations
- 2009-25 Alex van Ballegooij (CWI)  
"RAM: Array Database Management through Relational Mapping"

- 2009-26 Fernando Koch (UU)  
An Agent-Based Model for the Development of Intelligent Mobile Services
- 2009-27 Christian Glahn (OU)  
Contextual Support of social Engagement and Reflection on the Web
- 2009-28 Sander Evers (UT)  
Sensor Data Management with Probabilistic Models
- 2009-29 Stanislav Pokraev (UT)  
Model-Driven Semantic Integration of Service-Oriented Applications
- 2009-30 Marcin Zukowski (CWI)  
Balancing vectorized query execution with bandwidth-optimized storage
- 2009-31 Sofiya Katrenko (UVA)  
A Closer Look at Learning Relations from Text
- 2009-32 Rik Farenhorst (VU) and Remco de Boer (VU)  
Architectural Knowledge Management: Supporting Architects and Auditors
- 2009-33 Khiet Truong (UT)  
How Does Real Affect Affect Affect Recognition In Speech?
- 2009-34 Inge van de Weerd (UU)  
Advancing in Software Product Management: An Incremental Method Engineering Approach
- 2009-35 Wouter Koelewijn (UL)  
Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling
- 2009-36 Marco Kalz (OUN)  
Placement Support for Learners in Learning Networks
- 2009-37 Hendrik Drachsler (OUN)  
Navigation Support for Learners in Informal Learning Networks
- 2009-38 Riina Vuorikari (OU)  
Tags and self-organisation: a metadata ecology for learning resources in a multilingual context
- 2009-39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin)  
Service Substitution – A Behavioral Approach Based on Petri Nets
- 2009-40 Stephan Raaijmakers (UvT)  
Multinomial Language Learning: Investigations into the Geometry of Language
- 2009-41 Igor Berezhnyy (UvT)  
Digital Analysis of Paintings
- 2009-42 Toine Bogers  
Recommender Systems for Social Bookmarking
- 2009-43 Virginia Nunes Leal Franqueira (UT)  
Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients
- 2009-44 Roberto Santana Tapia (UT)  
Assessing Business-IT Alignment in Networked Organizations
- 2009-45 Jilles Vreeken (UU)  
Making Pattern Mining Useful
- 2009-46 Loredana Afanasiev (UvA)  
Querying XML: Benchmarks and Recursion
- 
- 2010
- 
- 2010-01 Matthijs van Leeuwen (UU)  
Patterns that Matter
- 2010-02 Ingo Wassink (UT)  
Work flows in Life Science
- 2010-03 Joost Geurts (CWI)  
A Document Engineering Model and Processing Framework for Multimedia documents
- 2010-04 Olga Kulyk (UT)  
Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments
- 2010-05 Claudia Hauff (UT)  
Predicting the Effectiveness of Queries and Retrieval Systems
- 2010-06 Sander Bakkes (UvT)  
Rapid Adaptation of Video Game AI
- 2010-07 Wim Fikkert (UT)

- Gesture interaction at a Distance
- 2010-08 Krzysztof Siewicz (UL)  
Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments
- 2010-09 Hugo Kielman (UL)  
A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging
- 2010-10 Rebecca Ong (UL)  
Mobile Communication and Protection of Children
- 2010-11 Adriaan Ter Mors (TUD)  
The world according to MARP: Multi-Agent Route Planning
- 2010-12 Susan van den Braak (UU)  
Sensemaking software for crime analysis
- 2010-13 Gianluigi Folino (RUN)  
High Performance Data Mining using Bio-inspired techniques
- 2010-14 Sander van Splunter (VU)  
Automated Web Service Reconfiguration
- 2010-15 Lianne Bodestaff (UT)  
Managing Dependency Relations in Inter-Organizational Models
- 2010-16 Sicco Verwer (TUD)  
Efficient Identification of Timed Automata, theory and practice
- 2010-17 Spyros Kotoulas (VU)  
Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications
- 2010-18 Charlotte Gerritsen (VU)  
Caught in the Act: Investigating Crime by Agent-Based Simulation
- 2010-19 Henriette Cramer (UvA)  
People's Responses to Autonomous and Adaptive Systems
- 2010-20 Ivo Swartjes (UT)  
Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative
- 2010-21 Harold van Heerde (UT)  
Privacy-aware data management by means of data degradation
- 2010-22 Michiel Hildebrand (CWI)  
End-user Support for Access to Heterogeneous Linked Data
- 2010-23 Bas Steunebrink (UU)  
The Logical Structure of Emotions
- 2010-24 Dmytro Tykhonov  
Designing Generic and Efficient Negotiation Strategies
- 2010-25 Zulfiqar Ali Memon (VU)  
Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective
- 2010-26 Ying Zhang (CWI)  
XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines
- 2010-27 Marten Voulon (UL)  
Automatisch contracteren
- 2010-28 Arne Koopman (UU)  
Characteristic Relational Patterns
- 2010-29 Stratos Idreos (CWI)  
Database Cracking: Towards Auto-tuning Database Kernels
- 2010-30 Marieke van Erp (UvT)  
Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval
- 2010-31 Victor de Boer (UVA)  
Ontology Enrichment from Heterogeneous Sources on the Web
- 2010-32 Marcel Hiel (UvT)  
An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems
- 2010-33 Robin Aly (UT)  
Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval
- 2010-34 Teduh Dirgahayu (UT)  
Interaction Design in Service Compositions
- 2010-35 Dolf Trieschnigg (UT)

- Proof of Concept: Concept-based Biomedical Information Retrieval
- 2010-36 Jose Janssen (OU)  
Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification
- 2010-37 Niels Lohmann (TUE)  
Correctness of services and their composition
- 2010-38 Dirk Fahland (TUE)  
From Scenarios to components
- 2010-39 Ghazanfar Farooq Siddiqui (VU)  
Integrative modeling of emotions in virtual agents
- 2010-40 Mark van Assem (VU)  
Converting and Integrating Vocabularies for the Semantic Web
- 2010-41 Guillaume Chaslot (UM)  
Monte-Carlo Tree Search
- 2010-42 Sybren de Kinderen (VU)  
Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach
- 2010-43 Peter van Kranenburg (UU)  
A Computational Approach to Content-Based Retrieval of Folk Song Melodies
- 2010-44 Pieter Bellekens (TUE)  
An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain
- 2010-45 Vasilios Andrikopoulos (UvT)  
A theory and model for the evolution of software services
- 2010-46 Vincent Pijpers (VU)  
e3alignment: Exploring Inter-Organizational Business-ICT Alignment
- 2010-47 Chen Li (UT)  
Mining Process Model Variants: Challenges, Techniques, Examples
- 2010-48 Milan Lovric (EUR)  
Behavioral Finance and Agent-Based Artificial Markets
- 2010-49 Jahn-Takeshi Saito (UM)  
Solving difficult game positions
- 2010-50 Bouke Huurnink (UVA)  
Search in Audiovisual Broadcast Archives
- 2010-51 Alia Khairia Amin (CWI)  
Understanding and supporting information seeking tasks in multiple sources
- 2010-52 Peter-Paul van Maanen (VU)  
Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention
- 2010-53 Edgar Meij (UVA)  
Combining Concepts and Language Models for Information Access

---

## 2011

---

- 2011-01 Botond Cseke (RUN)  
Variational Algorithms for Bayesian Inference in Latent Gaussian Models
- 2011-02 Nick Tinnemeier(UU)  
Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language
- 2011-03 Jan Martijn van der Werf (TUE)  
Compositional Design and Verification of Component-Based Information Systems
- 2011-04 Hado van Hasselt (UU)  
Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms
- 2011-05 Base van der Raadt (VU)  
Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.
- 2011-06 Yiwen Wang (TUE)  
Semantically-Enhanced Recommendations in Cultural Heritage
- 2011-07 Yujia Cao (UT)



- Multimodal Information Presentation for High Load Human Computer Interaction
- 2011-08 Nieske Vergunst (UU)  
BDI-based Generation of Robust Task-Oriented Dialogues
- 2011-09 Tim de Jong (OU)  
Contextualised Mobile Media for Learning
- 2011-10 Bart Bogaert (UvT)  
Cloud Content Contention
- 2011-11 Dhaval Vyas (UT)  
Designing for Awareness: An Experience-focused HCI Perspective
- 2011-12 Carmen Bratosin (TUE)  
Grid Architecture for Distributed Process Mining
- 2011-13 Xiaoyu Mao (UvT)  
Airport under Control. Multiagent Scheduling for Airport Ground Handling
- 2011-14 Milan Lovric (EUR)  
Behavioral Finance and Agent-Based Artificial Markets
- 2011-15 Marijn Koolen (UvA)  
The Meaning of Structure: the Value of Link Evidence for Information Retrieval
- 2011-16 Maarten Schadd (UM)  
Selective Search in Games of Different Complexity
- 2011-17 Jiyin He (UVA)  
Exploring Topic Structure: Coherence, Diversity and Relatedness
- 2011-18 Mark Ponsen (UM)  
Strategic Decision-Making in complex games
- 2011-19 Ellen Rusman (OU)  
The Mind's Eye on Personal Profiles
- 2011-20 Qing Gu (VU)  
Guiding service-oriented software engineering - A view-based approach
- 2011-21 Linda Terlouw (TUD)  
Modularization and Specification of Service-Oriented Systems
- 2011-22 Junte Zhang (UVA)  
System Evaluation of Archival Description and Access
- 2011-23 Wouter Weerkamp (UVA)  
Finding People and their Utterances in Social Media
- 2011-24 Herwin van Welbergen (UT)  
Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior
- 2011-25 Syed Waqar ul Qounain Jaffry (VU)  
Analysis and Validation of Models for Trust Dynamics
- 2011-26 Matthijs Aart Pontier (VU)  
Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots
- 2011-27 Aniel Bhulai (VU)  
Dynamic website optimization through autonomous management of design patterns
- 2011-28 Rianne Kaptein(UVA)  
Effective Focused Retrieval by Exploiting Query Context and Document Structure
- 2011-29 Faisal Kamiran (TUE)  
Discrimination-aware Classification
- 2011-30 Egon van den Broek (UT)  
Affective Signal Processing (ASP): Unraveling the mystery of emotions
- 2011-31 Ludo Waltman (EUR)  
Computational and Game-Theoretic Approaches for Modeling Bounded Rationality
- 2011-32 Nees-Jan van Eck (EUR)  
Methodological Advances in Bibliometric Mapping of Science
- 2011-33 Tom van der Weide (UU)  
Arguing to Motivate Decisions
- 2011-34 Paolo Turrini (UU)  
Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations

- 2011-35 Maaïke Harbers (UU)  
Explaining Agent Behavior in Virtual Training
- 2011-36 Erik van der Spek (UU)  
Experiments in serious game design: a cognitive approach
- 2011-37 Adriana Burlutiu (RUN)  
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference
- 2011-38 Nyree Lemmens (UM)  
Bee-inspired Distributed Optimization
- 2011-39 Joost Westra (UU)  
Organizing Adaptation using Agents in Serious Games
- 2011-40 Viktor Clerc (VU)  
Architectural Knowledge Management in Global Software Development
- 2011-41 Luan Ibraimi (UT)  
Cryptographically Enforced Distributed Data Access Control
- 2011-42 Michal Sindlar (UU)  
Explaining Behavior through Mental State Attribution
- 2011-43 Henk van der Schuur (UU)  
Process Improvement through Software Operation Knowledge
- 2011-44 Boris Reuderink (UT)  
Robust Brain-Computer Interfaces
- 2011-45 Herman Stehouwer (UvT)  
Statistical Language Models for Alternative Sequence Selection
- 2011-46 Beibei Hu (TUD)  
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work
- 2011-47 Azizi Bin Ab Aziz (VU)  
Exploring Computational Models for Intelligent Support of Persons with Depression
- 2011-48 Mark Ter Maat (UT)  
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent
- 2011-49 Andreea Niculescu (UT)  
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

---

2011

- 
- 2012-01 Terry Kakeeto (UvT)  
Relationship Marketing for SMEs in Uganda
  - 2012-02 Muhammad Umair (VU)  
Adaptivity, emotion, and Rationality in Human and Ambient Agent Models
  - 2012-03 Adam Vanya (VU)  
Supporting Architecture Evolution by Mining Software Repositories
  - 2012-04 Jurriaan Souer (UU)  
Development of Content Management System-based Web Applications
  - 2012-05 Marijn Plomp (UU)  
Maturing Interorganisational Information Systems
  - 2012-06 Wolfgang Reinhardt (OU)  
Awareness Support for Knowledge Workers in Research Networks
  - 2012-07 Rianne van Lambalgen (VU)  
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions
  - 2012-08 Gerben de Vries (UVA)  
Kernel Methods for Vessel Trajectories
  - 2012-09 Ricardo Neisse (UT)  
Trust and Privacy Management Support for Context-Aware Service Platforms
  - 2012-10 David Smits (TUE)  
Towards a Generic Distributed Adaptive Hypermedia Environment
  - 2012-11 J.C.B. Rantham Prabhakara (TUE)

- Process Mining in the Large: Preprocessing, Discovery, and Diagnostics
- 2012-12 Kees van der Sluijs (TUE)  
Model Driven Design and Data Integration in Semantic Web Information Systems
- 2012-13 Suleman Shahid (UvT)  
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions
- 2012-14 Evgeny Knutov (TUE)  
Generic Adaptation Framework for Unifying Adaptive Web-based Systems
- 2012-15 Natalie van der Wal (VU)  
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes
- 2012-16 Fiemke Both (VU)  
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment
- 2012-17 Amal Elgammal (UvT)  
Towards a Comprehensive Framework for Business Process Compliance
- 2012-18 Eltjo Poort (VU)  
Improving Solution Architecting Practices
- 2012-19 Helen Schonenberg (TUE)  
What's Next? Operational Support for Business Process Execution
- 2012-20 Ali Bahramisharif (RUN)  
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing
- 2012-21 Roberto Cornacchia (TUD)  
Querying Sparse Matrices for Information Retrieval
- 2012-22 Thijs Vis (UvT)  
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?
- 2012-23 Christian Muehl (UT)  
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction
- 2012-24 Laurens van der Werff (UT)  
Evaluation of Noisy Transcripts for Spoken Document Retrieval
- 2012-25 Silja Eckartz (UT)  
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application
- 2012-26 Emile de Maat (UVA)  
Making Sense of Legal Text
- 2012-27 Hayrettin Gurkok (UT)  
Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games
- 2012-28 Nancy Pascall (UvT)  
Engendering Technology Empowering Women
- 2012-29 Almer Tigelaar (UT)  
Peer-to-Peer Information Retrieval
- 2012-30 Alina Pommeranz (TUD)  
Designing Human-Centered Systems for Reflective Decision Making
- 2012-31 Emily Bagarukayo (RUN)  
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure
- 2012-32 Wietske Visser (TUD)  
Qualitative multi-criteria preference representation and reasoning
- 2012-33 Rory Sie (OUN)  
Coalitions in Cooperation Networks (COCOON)
- 2012-34 Pavol Jancura (RUN)  
Evolutionary analysis in PPI networks and applications
- 2012-35 Evert Haasdijk (VU)  
Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics
- 2012-36 Denis Ssebugwawo (RUN)  
Analysis and Evaluation of Collaborative Modeling Processes
- 2012-37 Agnes Nakakawa (RUN)

- A Collaboration Process for Enterprise Architecture Creation
- 2012-38 Selmar Smit (VU)  
Parameter Tuning and Scientific Testing in Evolutionary Algorithms
- 2012-39 Hassan Fatemi (UT)  
Risk-aware design of value and coordination networks
- 2012-40 Agus Gunawan (UvT)  
Information Access for SMEs in Indonesia
- 2012-41 Sebastian Kelle (OU)  
Game Design Patterns for Learning
- 2012-42 Dominique Verpoorten (OU)  
Reflection Amplifiers in self-regulated Learning
- 2012-43 *Withdrawn*
- 2012-44 Anna Tordai (VU)  
On Combining Alignment Techniques
- 2012-45 Benedikt Kratz (UvT)  
A Model and Language for Business-aware Transactions
- 2012-46 Simon Carter (UVA)  
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation
- 2012-47 Manos Tsagkias (UVA)  
Mining Social Media: Tracking Content and Predicting Behavior
- 2012-48 Jorn Bakker (TUE)  
Handling Abrupt Changes in Evolving Time-series Data
- 2012-49 Michael Kaisers (UM)  
Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions
- 2012-50 Steven van Kervel (TUD)  
Ontology driven Enterprise Information Systems Engineering
- 2012-51 Jeroen de Jong (TUD)  
Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching

---

2013

- 
- 2013-01 Viorel Milea (EUR)  
News Analytics for Financial Decision Support
- 2013-02 Erietta Liarou (CWI)  
MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing
- 2013-03 Szymon Klarman (VU)  
Reasoning with Contexts in Description Logics
- 2013-04 Chetan Yadati (TUD)  
Coordinating autonomous planning and scheduling
- 2013-05 Dulce Pumareja (UT)  
Groupware Requirements Evolutions Patterns
- 2013-06 Romulo Goncalves(CWI)  
The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience
- 2013-07 Giel van Lankveld (UvT)  
Quantifying Individual Player Differences
- 2013-08 Robbert-Jan Merk (VU)  
Making enemies: cognitive modeling for opponent agents in fighter pilot simulators
- 2013-09 Fabio Gori (RUN)  
Metagenomic Data Analysis: Computational Methods and Applications
- 2013-10 Jeewanie Jayasinghe Arachchige (UvT)  
A Unified Modeling Framework for Service Design
- 2013-11 Evangelos Pournaras (TUD)  
Multi-level Reconfigurable Self-organization in Overlay Services
- 2013-12 Marian Razavian (VU)  
Knowledge-driven Migration to Services
- 2013-13 Mohammad Safiri(UT)

- Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly
- 2013-14 Jafar Tanha (UVA)  
Ensemble Approaches to Semi-Supervised Learning Learning
- 2013-15 Daniel Hennes (UM)  
Multiagent Learning - Dynamic Games and Applications
- 2013-16 Eric Kok (UU)  
Exploring the practical benefits of argumentation in multi-agent deliberation
- 2013-17 Koen Kok (VU)  
The PowerMatcher: Smart Coordination for the Smart Electricity Grid
- 2013-18 Jeroen Janssens (UvT)  
Outlier Selection and One-Class Classification
- 2013-19 Renze Steenhuizen (TUD)  
Coordinated Multi-Agent Planning and Scheduling
- 2013-20 Katja Hofmann (UvA)  
Fast and Reliable Online Learning to Rank for Information Retrieval
- 2013-21 Sander Wubben (UvT)  
Text-to-text generation by monolingual machine translation
- 2013-22 Tom Claassen (RUN)  
Causal Discovery and Logic
- 2013-23 Patricio de Alencar Silva(UvT)  
Value Activity Monitoring
- 2013-24 Haitham Bou Ammar (UM)  
Automated Transfer in Reinforcement Learning
- 2013-25 Agnieszka Anna Latoszek-Berendsen (UM)  
Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System
- 2013-26 Alireza Zarghami (UT)  
Architectural Support for Dynamic Homecare Service Provisioning
- 2013-27 Mohammad Huq (UT)  
Inference-based Framework Managing Data Provenance
- 2013-28 Frans van der Sluis (UT)  
When Complexity becomes Interesting: An Inquiry into the Information eXperience
- 2013-29 Iwan de Kok (UT)  
Listening Heads
- 2013-30 Joyce Nakatumba (TUE)  
Resource-Aware Business Process Management: Analysis and Support
- 2013-31 Dinh Khoa Nguyen (UvT)  
Blueprint Model and Language for Engineering Cloud Applications
- 2013-32 Kamakshi Rajagopal (OUN)  
Networking For Learning: The role of Networking in a Lifelong Learner's Professional Development
- 2013-33 Qi Gao (TUD)  
User Modeling and Personalization in the Microblogging Sphere
- 2013-34 Kien Tjin-Kam-Jet (UT)  
Distributed Deep Web Search
- 2013-35 Abdallah El Ali (UvA)  
Minimal Mobile Human Computer Interaction
- 2013-36 Than Lam Hoang (TUE)  
Pattern Mining in Data Streams
- 2013-37 Dirk Börner (OUN)  
Ambient Learning Displays
- 2013-38 Eelco den Heijer (VU)  
Autonomous Evolutionary Art
- 2013-39 Joop de Jong (TUD)  
A Method for Enterprise Ontology based Design of Enterprise Information Systems
- 2013-40 Pim Nijssen (UM)  
Monte-Carlo Tree Search for Multi-Player Games

- 2013-41 Jochem Liem (UVA)  
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning
- 2013-42 Léon Planken (TUD)  
Algorithms for Simple Temporal Reasoning
- 2013-43 Marc Bron (UVA)  
Exploration and Contextualization through Interaction and Concepts
- 

2014

---

- 2014-01 Nicola Barile (UU)  
Studies in Learning Monotone Models from Data
- 2014-02 Fiona Tulyano (RUN)  
Combining System Dynamics with a Domain Modeling Method
- 2014-03 Sergio Raul Duarte Torres (UT)  
Information Retrieval for Children: Search Behavior and Solutions
- 2014-04 Hanna Jochmann-Mannak (UT)  
Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation
- 2014-05 Jurriaan van Reijssen (UU)  
Knowledge Perspectives on Advancing Dynamic Capability
- 2014-06 Damian Tamburri (VU)  
Supporting Networked Software Development
- 2014-07 Arya Adriansyah (TUE)  
Aligning Observed and Modeled Behavior
- 2014-08 Samur Araujo (TUD)  
Data Integration over Distributed and Heterogeneous Data Endpoints
- 2014-09 Philip Jackson (UvT)  
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language
- 2014-10 Ivan Salvador Razo Zapata (VU)  
Service Value Networks
- 2014-11 Janneke van der Zwaan (TUD)  
An Empathic Virtual Buddy for Social Support
- 2014-12 Willem van Willigen (VU)  
Look Ma, No Hands: Aspects of Autonomous Vehicle Control
- 2014-13 Arlette van Wissen (VU)  
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains
- 2014-14 Yangyang Shi (TUD)  
Language Models With Meta-information
- 2014-15 Natalya Mogles (VU)  
Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare
- 2014-16 Krystyna Milian (VU)  
Supporting trial recruitment and design by automatically interpreting eligibility criteria
- 2014-17 Kathrin Dentler (VU)  
Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability
- 2014-18 Mattijs Ghijsen (VU)  
Methods and Models for the Design and Study of Dynamic Agent Organizations
- 2014-19 Vincius Ramos (TUE)  
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support
- 2014-20 Mena Habib (UT)  
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link
- 2014-21 Cassidy Clark (TUD)  
Negotiation and Monitoring in Open Environments
- 2014-22 Marieke Peeters (UU)  
Personalized Educational Games - Developing agent-supported scenario-based training

- 2014-23 Eleftherios Sidirourgos (UvA/CWI)  
Space Efficient Indexes for the Big Data Era
- 2014-24 Davide Ceolin (VU)  
Trusting Semi-structured Web Data
- 2014-25 Martijn Lappenschaar (RUN)  
New network models for the analysis of disease interaction
- 2014-26 Tim Baarslag (TUD)  
What to Bid and When to Stop
- 2014-27 Rui Jorge Almeida (EUR)  
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty
- 2014-28 Anna Chmielowiec (VU)  
Decentralized k-Clique Matching
- 2014-29 Jaap Kabbedijk (UU)  
Variability in Multi-Tenant Enterprise Software
- 2014-30 Peter de Kock Berenschot (UvT)  
Anticipating Criminal Behaviour
- 2014-31 Leo van Moergestel (UU)  
Agent Technology in Agile Multiparallel Manufacturing and Product Support
- 2014-32 Naser Ayat (UVA)  
On Entity Resolution in Probabilistic Data
- 2014-33 Tesfa Tegegne Asfaw (RUN)  
Service Discovery in eHealth
- 2014-34 Christina Manteli (VU)  
The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems
- 2014-35 Joost van Oijen (UU)  
Cognitive Agents in Virtual Worlds: A Middleware Design Approach
- 2014-36 Joos Buijs (TUE)  
Flexible Evolutionary Algorithms for Mining Structured Process Models
- 2014-37 Maral Dadvar (UT)  
Experts and Machines United Against Cyberbullying
- 2014-38 Danny Plass-Oude Bos (UT)  
Making brain-computer interfaces better: improving usability through post-processing
- 2014-39 Jasmina Maric (UvT)  
Web Communities, Immigration and Social Capital
- 2014-40 Walter Oboma (RUN)  
A Framework for Knowledge Management Using ICT in Higher Education
- 2014-41 Frederik Hogenboom (EUR)  
Automated Detection of Financial Events in News Text
- 2014-42 Carsten Eickhoff (CWI/TUD)  
Contextual Multidimensional Relevance Models
- 2014-43 Kevin Vlaanderen (UU)  
Supporting Process Improvement using Method Increments
- 2014-44 Paulien Meesters (UvT)  
Intelligent Blauw. Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden
- 2014-45 Birgit Schmitz (OU)  
Mobile Games for Learning: A Pattern-Based Approach
- 2014-46 Ke Tao (TUD)  
Social Web Data Analytics: Relevance, Redundancy, Diversity
- 2014-47 Shangsong Liang (UVA)  
Fusion and Diversification in Information Retrieval

---

2015

- 
- 2015-01 Niels Netten (UVA)  
Machine Learning for Relevance of Information in Crisis Response
- 2015-02 Faiza Bukhsh (UvT)  
Smart auditing: Innovative Compliance Checking in Customs Controls
- 2015-03 Twan van Laarhoven (RUN)

- Machine learning for network data
- 2015-04 Howard Spoelstra (OUN)  
Collaborations in Open Learning Environments
- 2015-05 Christoph Bosch(UT)  
Cryptographically Enforced Search Pattern Hiding
- 2015-06 Farideh Heidari (TUD)  
Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes
- 2015-07 Maria-Hendrike Peetz(UvA)  
Time-Aware Online Reputation Analysis
- 2015-08 Jie Jiang (TUD)  
Organizational Compliance: An agent-based model for designing and evaluating organizational interactions
- 2015-09 Randy Klaassen(UT)  
HCI Perspectives on Behavior Change Support Systems
- 2015-10 Henry Hermans (OUN)  
OpenU: design of an integrated system to support lifelong learning
- 2015-11 Yongming Luo(TUE)  
Designing algorithms for big graph datasets: A study of computing bisimulation and joins
- 2015-12 Julie M. Birkholz (VU)  
Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks
- 2015-13 Giuseppe Procaccianti(VU)  
Energy-Efficient Software
- 2015-14 Bart van Straalen (UT)  
A cognitive approach to modeling bad news conversations
- 2015-15 Klaas Andries de Graaf (VU)  
Ontology-based Software Architecture Documentation
- 2015-16 Changyun Wei (UT)  
Cognitive Coordination for Cooperative Multi-Robot Teamwork
- 2015-17 Andre van Cleeff (UT)  
Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs
- 2015-18 Holger Pirk (CWI)  
Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories
- 2015-19 Bernardo Tabuenca (OUN)  
Ubiquitous Technology for Lifelong Learners
- 2015-20 Lois Vanhee(UU)  
Using Culture and Values to Support Flexible Coordination
- 2015-21 Sibren Fetter (OUN)  
Using Peer-Support to Expand and Stabilize Online Learning
- 2015-22 Zheming Zhu(UT)  
Co-occurrence Rate Networks
- 2015-23 Luit Gazendam (VU)  
Cataloguer Support in Cultural Heritage
- 2015-24 Richard Berendsen (UVA)  
Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation
- 2015-25 Steven Woudenberg (UU)  
Bayesian Tools for Early Disease Detection
- 2015-26 Alexander Hogenboom (EUR)  
Sentiment Analysis of Text Guided by Semantics and Structure
- 2015-27 Sandor Heman (CWI)  
Updating compressed column stores
- 2015-28 Janet Bagorogozo(TiU)  
Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO
- 2015-29 Hendrik Baier (UM)  
Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains
- 2015-30 Kiavash Bahreini(OU)  
Real-time Multimodal Emotion Recognition in E-Learning



- 2015-31 Yakup Koc (TUD)  
On the robustness of Power Grids
- 2015-32 Jerome Gard(UL)  
Corporate Venture Management in SMEs
- 2015-33 Frederik Schadd (TUD)  
Ontology Mapping with Auxiliary Resources
- 2015-34 Victor de Graaf(UT)  
Gesocial Recommender Systems
- 2015-35 Jungxao Xu (TUD)  
Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction
- 

2016

---

- 2016-01 Syed Saiden Abbas (RUN)  
Recognition of Shapes by Humans and Machines
- 2016-02 Michiel Christiaan Meulendijk (UU)  
Optimizing medication reviews through decision support: prescribing a better pill to swallow
- 2016-03 Maya Sappelli (RUN)  
Knowledge Work in Context: User Centered Knowledge Worker Support
- 2016-04 Laurens Rietveld (VU)  
Publishing and Consuming Linked Data
- 2016-05 Evgeny Sherkhonov (UVA)  
Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers
- 2016-06 Michel Wilson (TUD)  
Robust scheduling in an uncertain environment
- 2016-07 Jeroen de Man (VU)  
Measuring and modeling negative emotions for virtual training
- 2016-08 Matje van de Camp (TiU)  
A Link to the Past: Constructing Historical Social Networks from Unstructured Data
- 2016-09 Archana Nottamkandath (VU)  
Trusting Crowdsourced Information on Cultural Artefacts
- 2016-10 George Karafotias (VUA)  
Parameter Control for Evolutionary Algorithms
- 2016-11 Anne Schuth (UVA)  
Search Engines that Learn from Their Users
- 2016-12 Max Knobbout (UU)  
Logics for Modelling and Verifying Normative Multi-Agent Systems
- 2016-13 Nana Baah Gyan (VU)  
The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach
- 2016-14 Ravi Khadka (UU)  
Revisiting Legacy Software System Modernization
- 2016-15 Steffen Michels (RUN)  
Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments
- 2016-16 Guangliang Li (UVA)  
Socially Intelligent Autonomous Agents that Learn from Human Reward
- 2016-17 Berend Weel (VU)  
Towards Embodied Evolution of Robot Organisms
- 2016-18 Albert Merono Penuela  
Refining Statistical Data on the Web
- 2016-19 Julia Efremova (Tu/e)  
Mining Social Structures from Genealogical Data
- 2016-20 Daan Odijk (UVA)  
Context & Semantics in News & Web Search
- 2016-21 Alejandro Moreno Celleri (UT)  
From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground
- 2016-22 Grace Lewis (VU)  
Software Architecture Strategies for Cyber-Foraging Systems

- 2016-23 Fei Cai (UVA)  
Query Auto Completion in Information Retrieval
- 2016-24 Brend Wanders (UT)  
Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
- 2016-25 Julia Kiseleva (TU/e)  
Using Contextual Information to Understand Searching and Browsing Behavior
- 2016-26 Dilhan Thilakaratne (VU)  
In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
- 2016-27 Wen Li (TUD)  
Understanding Geo-spatial Information on Social Media
- 2016-28 Mingxin Zhang (TUD)  
Large-scale Agent-based Social Simulation - A study on epidemic prediction and control
- 2016-29 Nicolas Honing (TUD)  
Peak reduction in decentralised electricity systems -Markets and prices for flexible planning
- 2016-30 Ruud Mattheij (UvT)  
The Eyes Have It
- 2016-31 Mohammad Khelghati (UT)  
Deep web content monitoring
- 2016-32 Eelco Vriezekolk (UT)  
Assessing Telecommunication Service Availability Risks for Crisis Organisations
- 2016-33 Peter Bloem (UVA)  
Single Sample Statistics, exercises in learning from just one example
- 2016-34 Dennis Schunselaar (TUE)  
Title: Configurable Process Trees: Elicitation, Analysis, and Enactment
- 2016-35 Zhaochun Ren  
Monitoring Social Media: Summarization, Classification and Recommendation
- 2016-36 Daphne Karreman (UT)  
Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
- 2016-37 Giovanni Sileno (UvA)  
Aligning Law and Action - a conceptual and computational inquiry
- 2016-38 Andrea Minuto (UT)  
Materials that Matter - Smart Materials meet Art & Interaction Design
- 2016-39 Merijn Bruijnes (UT)  
Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect
- 2016-40 Christian Detweiler (TUD)  
Accounting for Values in Design
- 2016-41 Thomas King (TUD)  
Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance
- 2016-42 Spyros Martzoukos (UVA)  
Combinatorial and Compositional Aspects of Bilingual Aligned Corpora